CSF tau and the CSF tau/ABeta ratio for the diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI)

Cochrane Library

Cochrane Database of Systematic Reviews

CSF tau and the CSF tau/ABeta ratio for the diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI) (Review)

Ritchie C, Smailagic N, Noel-Storr AH, Ukoumunne O, Ladds EC, Martin S

Ritchie C, Smailagic N, Noel-Storr AH, Ukoumunne O, Ladds EC, Martin S.
CSF tau and the CSF tau/ABeta ratio for the diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI).
Cochrane Database of Systematic Reviews 2017, Issue 3. Art. No.: CD010803.
DOI: 10.1002/14651858.CD010803.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Review]

CSF tau and the CSF tau/ABeta ratio for the diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI)

Craig Ritchie[1], Nadja Smailagic[2a], Anna H Noel-Storr[3], Obioha Ukoumunne[4], Emma C Ladds[5], Steven Martin[2]

[1]Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK. [2]Institute of Public Health, University of Cambridge, Cambridge, UK. [3]Radcliffe Department of Medicine, University of Oxford, Oxford, UK. [4]NIHR CLAHRC South West Peninsula (PenCLAHRC), University of Exeter Medical School, University of Exeter, Exeter, UK. [5]Southmead hospital, North Bristol NHS Trust, Bristol, UK

[a]Nadja Smailagic is joint first author for this review

Contact address: Craig Ritchie, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK. craig.ritchie@ed.ac.uk.

Editorial group: Cochrane Dementia and Cognitive Improvement Group.
Publication status and date: New, published in Issue 3, 2017.

Citation: Ritchie C, Smailagic N, Noel-Storr AH, Ukoumunne O, Ladds EC, Martin S. CSF tau and the CSF tau/ABeta ratio for the diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database of Systematic Reviews* 2017, Issue 3. Art. No.: CD010803. DOI: 10.1002/14651858.CD010803.pub2.

ABSTRACT

Background

Research suggests that measurable change in cerebrospinal fluid (CSF) biomarkers occurs years in advance of the onset of clinical symptoms (Beckett 2010). In this review, we aimed to assess the ability of CSF tau biomarkers (t-tau and p-tau) and the CSF tau (t-tau or p-tau)/ABeta ratio to enable the detection of Alzheimer's disease pathology in patients with mild cognitive impairment (MCI). These biomarkers have been proposed as important in new criteria for Alzheimer's disease dementia that incorporate biomarker abnormalities.

Objectives

To determine the diagnostic accuracy of 1) CSF t-tau, 2) CSF p-tau, 3) the CSF t-tau/ABeta ratio and 4) the CSF p-tau/ABeta ratio index tests for detecting people with MCI at baseline who would clinically convert to Alzheimer's disease dementia or other forms of dementia at follow-up.

Search methods

The most recent search for this review was performed in January 2013. We searched MEDLINE (OvidSP), Embase (OvidSP), BIOSIS Previews (Thomson Reuters Web of Science), Web of Science Core Collection, including Conference Proceedings Citation Index (Thomson Reuters Web of Science), PsycINFO (OvidSP), and LILACS (BIREME). We searched specialized sources of diagnostic test accuracy studies and reviews. We checked reference lists of relevant studies and reviews for additional studies. We contacted researchers for possible relevant but unpublished data. We did not apply any language or data restriction to the electronic searches. We did not use any methodological filters as a method to restrict the search overall.

Selection criteria

We selected those studies that had prospectively well-defined cohorts with any accepted definition of MCI and with CSF t-tau or p-tau and CSF tau (t-tau or p-tau)/ABeta ratio values, documented at or around the time the MCI diagnosis was made. We also included studies which looked at data from those cohorts retrospectively, and which contained sufficient data to construct two by two

tables expressing those biomarker results by disease status. Moreover, studies were only selected if they applied a reference standard for Alzheimer's disease dementia diagnosis, for example, the NINCDS-ADRDA or Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) criteria.

Data collection and analysis

We screened all titles generated by the electronic database searches. Two review authors independently assessed the abstracts of all potentially relevant studies, and the full papers for eligibility. Two independent assessors performed data extraction and quality assessment. Where data allowed, we derived estimates of sensitivity at fixed values of specificity from the model we fitted to produce the summary receiver operating characteristic (ROC) curve.

Main results

In total, 1282 participants with MCI at baseline were identified in the 15 included studies of which 1172 had analysable data; 430 participants converted to Alzheimer's disease dementia and 130 participants to other forms of dementia. Follow-up ranged from less than one year to over four years for some participants, but in the majority of studies was in the range one to three years.

Conversion to Alzheimer's disease dementia

The accuracy of the CSF t-tau was evaluated in seven studies (291 cases and 418 non-cases).The sensitivity values ranged from 51% to 90% while the specificity values ranged from 48% to 88%. At the median specificity of 72%, the estimated sensitivity was 75% (95% CI 67 to 85), the positive likelihood ratio was 2.72 (95% CI 2.43 to 3.04), and the negative likelihood ratio was 0.32 (95% CI 0.22 to 0.47).

Six studies (164 cases and 328 non-cases) evaluated the accuracy of the CSF p-tau. The sensitivities were between 40% and 100% while the specificities were between 22% and 86%. At the median specificity of 47.5%, the estimated sensitivity was 81% (95% CI: 64 to 91), the positive likelihood ratio was 1.55 (CI 1.31 to 1.84), and the negative likelihood ratio was 0.39 (CI: 0.19 to 0.82).

Five studies (140 cases and 293 non-cases) evaluated the accuracy of the CSF p-tau/ABeta ratio. The sensitivities were between 80% and 96% while the specificities were between 33% and 95%. We did not conduct a meta-analysis because the studies were few and small. Only one study reported the accuracy of CSF t-tau/ABeta ratio.

Our findings are based on studies with poor reporting. A significant number of studies had unclear risk of bias for the reference standard, participant selection and flow and timing domains. According to the assessment of index test domain, eight of 15 studies were of poor methodological quality.

The accuracy of these CSF biomarkers for 'other dementias' had not been investigated in the included primary studies.

Investigation of heterogeneity

The main sources of heterogeneity were thought likely to be reference standards used for the target disorders, sources of recruitment, participant sampling, index test methodology and aspects of study quality (particularly, inadequate blinding).

We were not able to formally assess the effect of each potential source of heterogeneity as planned, due to the small number of studies available to be included.

Authors' conclusions

The insufficiency and heterogeneity of research to date primarily leads to a state of uncertainty regarding the value of CSF testing of t-tau, p-tau or p-tau/ABeta ratio for the diagnosis of Alzheimer's disease in current clinical practice. Particular attention should be paid to the risk of misdiagnosis and overdiagnosis of dementia (and therefore over-treatment) in clinical practice. These tests, like other biomarker tests which have been subject to Cochrane DTA reviews, appear to have better sensitivity than specificity and therefore might have greater utility in ruling out Alzheimer's disease as the aetiology to the individual's evident cognitive impairment, as opposed to ruling it in. The heterogeneity observed in the few studies awaiting classification suggests our initial summary will remain valid. However, these tests may have limited clinical value until uncertainties have been addressed. Future studies with more uniformed approaches to thresholds, analysis and study conduct may provide a more homogenous estimate than the one that has been available from the included studies we have identified.

PLAIN LANGUAGE SUMMARY

Proteins in cerebrospinal fluids (CSF) for early prediction of developing Alzheimer's disease or other dementia in people with mild cognitive problems

Background

The numbers of people with dementia and other cognitive problems are increasing globally. A diagnosis of dementia at early stage is recommended but there is no agreement on the best approach. A range of tests have been developed which healthcare professionals can use to assess people with poor memory or cognitive impairment. In this review, we have focused on the cerebrospinal fluid (CSF) diagnostic tests.

Review question

We reviewed the evidence about the accuracy of CSF tests in identifying those people presenting with mild cognitive impairment (MCI) who would develop Alzheimer's disease dementia or other forms of dementia over a period of time.

Study characteristics

The evidence is current to January 2013. We included 15 studies containing a total of 1282 participants with MCI. The majority of studies (n = 9) were published between 2010 and 2013. The remaining six studies were published between 2004 and 2009. All of the included studies were conducted in Europe.

Study sizes varied and ranged from 15 to 231 participants.The mean (range) age of the youngest sample was 64 years (45 to 76) and the mean (standard deviation) age of the oldest sample was 73.4 (6.6) years.

Quality of the evidence

Our findings are based on studies with poor reporting, with a majority of studies at unclear risk of bias due to insufficient details given on how participants were selected and how the clinical diagnosis of dementia was established. According to the assessment of how the CSF tests were conducted and analysed, eight of 15 studies were of poor methodological quality.

Key findings

Below is a summary of key findings for the tests:

CSF t-tau test for conversion from MCI to Alzheimer's disease dementia

The sensitivity values in seven individual studies ranged from 51% to 90% while the specificity values ranged from 48% to 88%. The statistical analysis of those studies showed that, at the fixed specificity of 72%, the estimated sensitivity was 77%, and, at the prevalence of 37%, the positive predictive value was 62% and the negative predictive value was 84%. Based on these results, on average 62 out of 100 people with MCI and a positive index test result would convert to Alzheimer's disease dementia but 38 would not; on average, 84 out of 100 people with MCI and with a negative index test result would not convert to Alzheimer's disease dementia, but 16 would.

CSF p-tau test for conversion from MCI to Alzheimer's disease dementia

The sensitivity values in six individual studies ranged from 40% to 100% while the specificity values ranged from 22% to 86%. The statistical analysis of those studies showed that, at the fixed specificity of 48%, the estimated sensitivity was 81%, and, at the prevalence of 37%, the positive predictive value was 48% and the negative predictive value was 81%. Based on these results, on average 48 out of 100 people with MCI and with a positive index test result would convert to Alzheimer's disease dementia, but 52 would not; on average, 81 out of 100 people with MCI with a negative index test result would not convert to Alzheimer's disease dementia, but 19 would.

We found that the cerebrospinal fluid (CSF) diagnostic test, as a single test, lacks the accuracy to identify those people with mild cognitive impairment (MCI) who would develop Alzheimer's disease dementia or other forms of dementia over a period of time. The data suggested that a negative CSF test, in people with MCI, almost indicates the absence of Alzheimer's disease as the cause of their clinical symptoms. However, a positive CSF test does not confirm the presence of Alzheimer's disease as the aetiology (cause) of their clinical symptoms.

There were methodological problems in the included studies that did not allow for a clear answer to the review question. The main limitations of the review were poor reporting in the included studies, lack of a widely accepted threshold of the CSF diagnostic tests in people with MCI, variability in length of follow-up, and the marked variation in CSF tests' accuracy between the included studies.

BACKGROUND

Dementia is a progressive syndrome of global cognitive impairment with resultant functional decline. In the United Kingdom (UK), it affects 5% of the population over 65 and 25% of those over 85 (Knapp 2007). Worldwide, there were estimated to be 36 million people living with dementia in 2010 (Wilmo 2010), and this will increase to over 115 million by 2050 (Prince 2013). The greatest increases in prevalence are likely to be seen in the developing regions. By 2040, China and its western-Pacific neighbours are predicted to have 26 million people living with dementia (Ferri 2005).

Dementia encompasses a group of neurodegenerative disorders that are characterised by progressive loss of cognitive function and ability to perform activities of daily living, that can be accompanied by neuropsychiatric symptoms and challenging behaviours of varying type and severity. The underlying pathology is usually degenerative and subtypes of dementia include Alzheimer's disease dementia, vascular dementia, dementia with Lewy bodies, and frontotemporal dementia. There may be considerable overlap in the clinical and pathological presentations (MRC CFAS 2001), and there is often coexistence of Alzheimer's disease dementia, vascular dementia and other causes of neuronal atrophy (Matthews 2009; Savva 2009).

Alzheimer's disease dementia is an incurable, progressive, neurodegenerative condition which accounts for over 50% of all dementias, afflicting 5% of men and 6% of women over the age of 60 worldwide (World Health Organization 2010). Its prevalence increases exponentially with age, with Alzheimer's dementia affecting fewer than 1% of people aged from 60 to 64 years, but 24% to 33% of those over the age of 85 (Ferri 2005).

There have been over a dozen different definitions used to describe cognitive impairment that is somehow qualitatively different from so-called 'normal' ageing. The first complaints in people with Alzheimer's disease spectrum are often cognitive problems such as problems with planning and judgement, as well as the more characteristic memory complaints. This may lead to a diagnosis of Mild Cognitive Impairment (MCI) if formal testing reveals objective evidence of cognitive impairment. It has not been previously mandated which psychometric tests should be used to objectively define cognitive impairment. However, the objectivity of the cognitive impairment diagnosis is critical, as it differentiates this population from a group with subjective cognitive impairment, which is more likely to have a non-neurodegenerative aetiology. MCI is a heterogeneous condition, the diagnosis of which holds very little prognostic significance. There are four outcomes for those within an MCI population: progression to Alzheimer's disease dementia, progression to another dementia, maintaining stable MCI, and recovery. Currently, 16 different classifications are used to define MCI (Matthews 2008). In this protocol, MCI refers to this extended definition of MCI or to the clinical criteria defined by Petersen criteria or revised Petersen criteria (Petersen 1999; Petersen 2004; Winbald 2004) or to the Cognitive Dementia Rating (CDR = 0.5) scale (Morris 1993).

Studies indicate that an annual average of 5% to 15% of people with MCI progress to Alzheimer's disease dementia (Petersen 1999; Bruscoli 2004; Mattson 2009; Petersen 2009). This all depends on clinical profile, settings and investigation for vascular disease. At the present time, there is no clinical method to determine accurately which of those people with MCI will develop Alzheimer's disease dementia or other forms of dementia.

Recent consensus guidelines have been developed, e.g. the second iteration of International Working Group (IWG2) on 'prodromal dementia', which seeks to improve prognostic accuracy in the prodromal phase of Azheimer's dementia by the incorporation in criteria of Alzheimer's disease-related biomarkers (Dubois 2014). It is in this context, that reviews such as this one become especially relevant and timely.

Research suggests that measurable change in proton emission tomography (PET), magnetic resonance (MRI) and cerebrospinal fluid (CSF) biomarkers occurs years in advance of the onset of clinical symptoms (Beckett 2010). In this review, we aimed to assess the ability of CSF total tau (t-tau), CSF phosphorylated tau (p-tau), the CSF t-tau/ABeta ratio, and the CSF p-tau/ABeta ratio, to enable the detection of Alzheimer's dementia and other forms of dementia in people with MCI. These biomarkers have been chosen as they are considered to be the most intimately expressed biomarkers of the Alzheimer's disease core pathology; namely, the aggregation and fibrilisation of the amyloid plaque and hyperphosphorylation of tau. Consequentially, these biomarkers have been proposed as important in new criteria for Alzheimer's disease dementia that incorporate biomarker abnormalities. PET imaging of amyloid is now approved by both the FDA and EMA to rule out Alzheimer's disease as the aetiology of MCI, especially in individuals with unusual clinical presentations. However, manufacturers of these tracers have ongoing 'appropriate use criteria' ongoing post-marketing studies to learn where these tests have greatest usage and utility for the person's accurate diagnosis. Recent improvements to CSF sampling and the relatively inexpensive nature of this test compared with PET scanning means that it will remain the test of choice for documenting CSF protein abnormalities in neurodegenerative disease. Side effects are increasingly rare but include headache and local reactions at the site of the lumbar puncture. Patients on anticoagulative therapies (except aspirin) are considered at too high a risk by most practitioners to undergo this procedure for the diagnosis of Alzheimer's dementia.

Target condition being diagnosed

In this review, there are two target conditions: i) Alzheimer's disease dementia and ii) other forms of dementia, both of which were

assessed at follow-up.

We compared the index test results obtained at baseline with the results of the reference standard (clinical criteria) obtained at follow-up (delayed verification of clinical diagnosis).

Index test(s)

This review is part of a suite of reviews for assessing the accuracy of CSF ABeta (Ritchie 2014), PET Amyloid (Zhang 2014; Smailagic 2015), MMSE (Arevalo-Rodriguez 2015), and other index tests in identifying those people with MCI without clinical onset of dementia, who would develop Alzheimer's disease dementia or other forms of dementia during follow-up. We planned to consider the following:

Total tau (t-tau) and phosphorylated tau (p-tau) CSF biomarker tests

Tau is a microtubule-associated protein located primarily in neuronal axons. There are six different human isoforms, each of which has multiple phosphorylation sites. Physiologically tau interacts with tubulin and plays an important role in the organisation and stabilisation of microtubules. Independent of phosphorylation status, slightly increased levels of CSF total tau (t-tau) have been associated with ageing, vascular dementia, multiple sclerosis, AIDS dementia, head injury and tauopathy; significant increases with Creutzfeldt-Jakob disease and meningoencephalitis; and a threefold increase has been seen in Alzheimer's disease compared to normal controls (Shoji 2002). A systematic review of CSF biomarkers for Alzheimer's disease analysing 41 studies of CSF t-tau, demonstrated a specificity of 90% and sensitivity of 81% in diagnosing the condition (Blennow 2003).

The p-tau protein also has a number of potential phosphorylation sites (Billingsley 1997) and abnormal hyperphosphorylation has been shown to be associated with microtubule disruption and the formation of neurofibrillary tangles, dystrophic neurites surrounded by neuritic plaques, and neuropil threads, major components of Alzheimer's disease pathophysiology (see Mandelkow 1998). A systematic review of 11 studies of CSF p-tau in Alzheimer's disease indicated a diagnostic specificity and sensitivity of 92% and 80% respectively (Blennow 2003).

There is great interest around the use of biomarkers and imaging techniques for the prediction of progression from MCI populations to Alzheimer's disease dementia and other forms of dementia. The international consortium study Alzheimer Disease Neuroimaging Initiative (ADNI), performed between 2004 and 2009, has so far been a key cohort study for predicting the progression from MCI to Alzheimer's disease using biomarkers, and demonstrated a sensitivity and specificity of CSF t-tau of 70% and 92% and CSF p-Tau181 of 68% and 73% respectively (Petersen 2010).

T-tau/ABeta ratio and p-tau/ABeta ratio CSF biomarker tests

ABeta is produced mainly by neurons, secreted into the CSF and then cleared through the blood-brain barrier and degraded by the reticuloendothelial system. ABeta levels are thus regulated in strict equilibrium between the brain, CSF and blood (Shoji 1992), but, in Alzheimer's disease patients, ABeta42 forms insoluble amyloid and accumulates as intracerebral fibrils, resulting in decreased levels of CSF ABeta42 (Shoji 2001).

ABeta in CSF has only modest potential as a test for delayed verification of Alzheimer's disease (Ritchie 2014), with meta-analysis of studies being hampered by poor methodological quality (Noel-Storr 2013) and multiple thresholds being reported between studies (Ritchie 2011).

In 2001, the American Academy of Neurology produced practical guidelines for dementia, including three Class II or III reports in a systematic review of a combination study of ABeta42 and t-tau CSF levels. The sensitivity and specificity for diagnosis of Alzheimer's disease were 85% and 87% (Knopman 2001), supported by the 2001 systematic review revealing 83% to 100% sensitivity and 85% to 95% specificity for the CSF ABeta42 and t-tau combination assay (Blennow 2003). Again, the ADNI cohort study demonstrated that the t-tau/ABeta42 ratio could be used to predict conversion from MCI to Alzheimer's disease dementia, revealing a sensitivity of 86% and specificity of 85% (Petersen 2010).

Clinical pathway

Dementia develops over several years and there is a presumed period when people are asymptomatic, although disease pathology may have accumulated. Individuals or their relatives may first notice subtle impairments of short-term memory when the completion of complex tasks such as management of finances or medications becomes increasingly difficult. In the UK, people usually present to their general practitioner who may then refer them to a specialist following a brief cognitive test, clinical examination and exclusion of relevant physical illness. The biomarkers may then be administered by a specialist. There is, however, much regional variability in this, with Spain and Nordic countries favouring CSF sampling in their routine clinical work-up, whereas other countries, such as the UK, do not. However, many people with dementia do not present until much later in the disorder and they will, therefore, follow a different pathway to diagnosis, for example, being identified during an admission to general hospital for a physical illness. Thus, the pathway influences the accuracy of the diagnostic test. The accuracy of the test will vary with the experience of the administrator, and the accuracy of the subsequent diagnosis will vary with the history of referrals to the particular healthcare setting. Diagnostic assessment pathways may vary in other countries and diagnoses may be made by a variety of specialists including psychiatrists, neurologists, and geriatricians.

Role of index test(s)

The sampling of CSF and assay for levels of tau and ABeta could have a role when applied in specialist clinics. Due to the costs,

risks, and complexity of the testing, CSF tests will not be applied in a primary care setting. The roles of these index tests are as add-on biomarker tests which have been proposed in new research diagnostic criteria to compliment clinical examination and cognitive tests.

Alternative test(s)

We did not include alternative tests in this review, because there are currently no standard practice tests available for the diagnosis of dementia.

Rationale

Recently proposed research diagnostic criteria for 'prodromal dementia'/'pre-dementia stage'/'MCI due to Alzheimer's disease pathology' and for 'Alzheimer's disease' and for the 'preclinical states of Alzheimer's disease' (Albert 2011; Dubois 2010; Dubois 2014), incorporate biomarkers based on imaging or CSF measures within the diagnostic rubric. These tests are core to the criteria, assuming they will improve the specificity of the traditional solely clinical criteria. It is crucial that each of these biomarkers is assessed for their diagnostic accuracy before they are adopted as routine tests in clinical practice. It is worth noting that in each of these criteria, a single abnormality in any of the proposed biomarker/imaging tests is considered sufficient to make a diagnosis of prodromal Alzheimer's disease dementia.

Underpinning the new criteria is the assumption that if Alzheimer's disease pathology can be diagnosed at an earlier, pre-dementia stage, this could open critical windows for interventions that will have a greater likelihood of success in affecting disease pathways and thereby improving clinical symptoms. Earlier accurate diagnosis will also help people with pre-dementia cognitive impairment, their families and potential carers make timely plans for the future. Coupled with appropriate contingency planning, proper recognition of the disease may also help to prevent inappropriate and potentially harmful admissions to hospital or institutional care (Bourne 2007). In addition, the accurate early identification of a dementia syndrome may improve opportunities for the use of newly evolving interventions designed to delay or prevent progression to more debilitating stages of dementia.

OBJECTIVES

To determine the diagnostic accuracy of 1) CSF t-tau, 2) CSF p-tau, 3) the CSF t-tau/ABeta ratio and 4) the CSF p-tau/ABeta ratio index tests for detecting people with MCI at baseline who would clinically convert to Alzheimer's disease dementia or other forms of dementia at follow-up.

Secondary objectives

To investigate the amount and associations of heterogeneity in the included studies of test accuracy.

We expected heterogeneity to be an important component of the review. We planned to use target population, index test, target disorder and study quality as a framework for the investigation of heterogeneity.

METHODS

Criteria for considering studies for this review

Types of studies

We considered longitudinal cohort studies in which index test results were obtained at baseline and the reference standard results at follow-up (see Index tests; Reference standards). These studies necessarily employ delayed verification of conversion to dementia and are sometimes labelled as 'delayed verification cross-sectional studies' (Bossuyt 2008,; Knottnerus 2002). This approach recognises the challenges of concurrent application of the reference test and index test. In reality, the reference standard for dementia is tissue sampling and histological examination, either at post mortem or from brain biopsy. Brain biopsy is not undertaken in any setting and a post mortem is so distant an event from the index test being conducted that there is the possibility that disease may have developed in the years after the index test. The Dementia DTA group chose to use later diagnosis of dementia (using standardised criteria) as evidence of delayed verification. This methodology has been published by our group (Mason 2010) and also reflects the approach taken in most of the primary research in this area.

We included nested case-control studies if they incorporated a delayed verification design. We believe this can only occur in the context of a cohort study, so these studies are invariably diagnostic nested cohort studies. W*e* only included data on performance of the index test to discriminate between people with MCI who converted to dementia and those who remained stable from those studies. We did not consider data from healthy controls or any other control group.

Participants

Participants recruited and clinically classified as those with mild cognitive impairment (MCI) at baseline were eligible for inclusion in this review. The diagnosis for MCI was established using the Petersen criteria or revised Petersen criteria (Petersen 1999; Petersen 2004; Winbald 2004) and/or Matthews criteria (Matthews 2008)

and/or 'CDR = 0.5' (Morris 1993). These criteria include: subjective complaints; a decline in memory objectively verified by neuropsychological testing in combination with a history from the patient; a decline in other cognitive domains; no or minimal impairment of activities of daily living; and not meeting the criteria for dementia. Therefore, the eligible participants had a number of tests, e.g. neuropsychological tests for cognitive deficit and checklists for activities of daily living, before study entry. Participants were defined either as amnestic single domain, amnestic multiple domain, non-amnestic single domain, non-amnestic multiple domain, or nonspecified MCI participants.

We included participants from secondary and tertiary settings. Although demographic and clinical characteristics of MCI, as well as sources of recruitment, might differ in those settings, we decided not to limit our review by setting; instead, we planned to look for variation within and between settings, and examined the potential influence of the setting on diagnostic performance of the index test in the analyses.

We excluded those studies that included people with MCI possibly caused by: i) a current or history of alcohol/drug abuse; ii) central nervous system (CNS) trauma (e.g. subdural haematoma), tumour, or infection; iii) other neurological conditions, e.g. Parkinson's or Huntington's diseases.

Because detail of the causes of study dropouts is crucial, and, if such data are missing, the reliability of the conclusions must be questioned, we planned to take this into consideration.

Index tests

Studies that assessed the accuracy of CSF measurements of CSF t-tau, CSF p-tau, CSF t-tau/ABeta ratio, or CSF p-tau/ABeta ratio were included.

There are currently no generally accepted standards for the plasma or CSF ABeta test threshold, and therefore it was not possible to prespecify what constituted a positive or negative result. We used the criteria which were applied in each included primary study to classify participants as either test positive or test negative.

Measure of index test: t-tau and p-tau and ABeta level in CSF (ng.l^{-1} or pg.ml^{-1})

The assays most commonly used were conventional Innogenetics, Ghent, Belgium kit or INNOTEST Phospho-Tau$_{(181)}$ kit or INNOTEST ABeta$_{42}$ or INNOTEST the multiplexing INNO-BIA AlzBio3 for CSF.

We did not include a comparator test because there are currently no standard practice tests available for the diagnosis of dementia. We compared the index tests with a reference standard.

Target conditions

There were two target conditions in this review:

1. Alzheimer's disease dementia (conversion from MCI to Alzheimer's disease dementia)
2. Any other forms of dementia (conversion from MCI to any other forms of dementia)

Reference standards

For the purpose of this review, several definitions of Alzheimer's disease dementia were acceptable. Included studies could apply probable or possible NINCDS-ADRDA (National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association) criteria (McKhann 1984). The Diagnostic and Statistical Manual of Mental Disorders (DSM) (DSMIII 1987; DSMIV 1994) and International Classification of Diseases (ICD) (World Health Organization 2010) definitions for Alzheimer's disease dementia were also acceptable. It should be noted that different iterations of these standards may not be directly comparable over time (e.g. DSM-IIIR versus DSM-IV). Moreover, the validity of the diagnoses may vary with the degree or manner in which the criteria have been operationalised (e.g. individual clinician versus algorithm versus consensus determination). We planned to consider all these issues in interpreting the results, using sensitivity analyses as appropriate.

Similarly, differing clinical definitions of other forms of dementias were acceptable. For Lewy body dementia, the reference standard is the McKeith criteria (McKeith 1996; McKeith 2005). For frontotemporal dementia, the reference standard is the Lund criteria (LMG 1994, Neary 1998, Boxer 2005). DSM (DSMIII 1987; DSMIV 1994) and ICD (World Health Organization 2010) were also acceptable for frontotemporal and vascular dementias.

The time interval over which progression from MCI to Alzheimer's disease dementia or other forms of dementia happened is also important. As age is the principal risk factor for Alzheimer's dementia and other forms of dementias, the longer the duration of follow-up, the more likely the possibility of generating false positive findings for the index test. To this end, no limits were put on the length of follow-up in the included studies, though this important variable was captured so we could examine between-study variations. This change reflected an alteration to the original thinking in the published protocol and is noted in the Differences between protocol and review section of this review.

We planned to segment analyses into separate follow-up periods for the delay in verification: less than one year, one year to less than two years; two to less than four years; and more than four years.

Search methods for identification of studies

Electronic searches

The main search for this review was performed in January 2013. However, we ran a top-up search in December 2015.

We searched MEDLINE (OvidSP), Embase (OvidSP), BIOSIS Previews (Thomson Reuters Web of Science), Web of Science Core Collection, including Conference Proceedings Citation Index (Thomson Reuters Web of Science), PsycINFO (OvidSP), and LILACS (BIREME) (see Appendix 1 for details of the sources searched, the search strategies used, and the number of hits that were retrieved for the search carried out in January 2013). The results of the top-up search that were carried out in December 2015 have not yet been fully incorporated into the review (please see Results of the search for more details).

We also requested a search of the Cochrane Register of Diagnostic Test Accuracy Studies (managed by the Cochrane Renal Group).

We did not apply any language or date restrictions to the electronic searches. We did not use methodological search filters (collections of terms aimed at reducing the number needed to screen by filtering out irrelevant records and retaining only those that are relevant) in the main bibliographic databases (MEDLINE, Embase and PsycINFO) as a single-stranded method to restrict the search overall because available filters have not yet proved sensitive enough for systematic review searches (Beynon 2013). Instead, we used a multi-stranded approach in order to maximise sensitivity, including some searches run in parallel, that included specific terms designed to capture diagnostic studies (see search narrative in Appendix 1)

Searching other resources

We checked the reference lists of all relevant studies for additional studies. We also conducted searches in the MEDION database (Meta-analyses van Diagnostisch Onderzoek) at www.mediondatabase.nl, Database of Abstracts of Reviews of Effects (DARE) at http://www.crd.york.ac.uk/CRDWeb, Health Technology Assessment Database (HTA Database) at http://www.crd.york.ac.uk/CRDWeb, and Aggressive Research Intelligence Facility (ARIF) database at www.arif.bham.ac.uk for other related systematic diagnostic accuracy reviews; we searched for systematic reviews of diagnostic studies from the International Federation of Clinical Chemistry and Laboratory Medicine Committee for Evidence-based Laboratory Medicine database (C-EBLM). We checked reference lists of any relevant systematic reviews for additional studies. We also contacted researchers involved in relevant studies for applicable and usable but unpublished data.

Data collection and analysis

Selection of studies

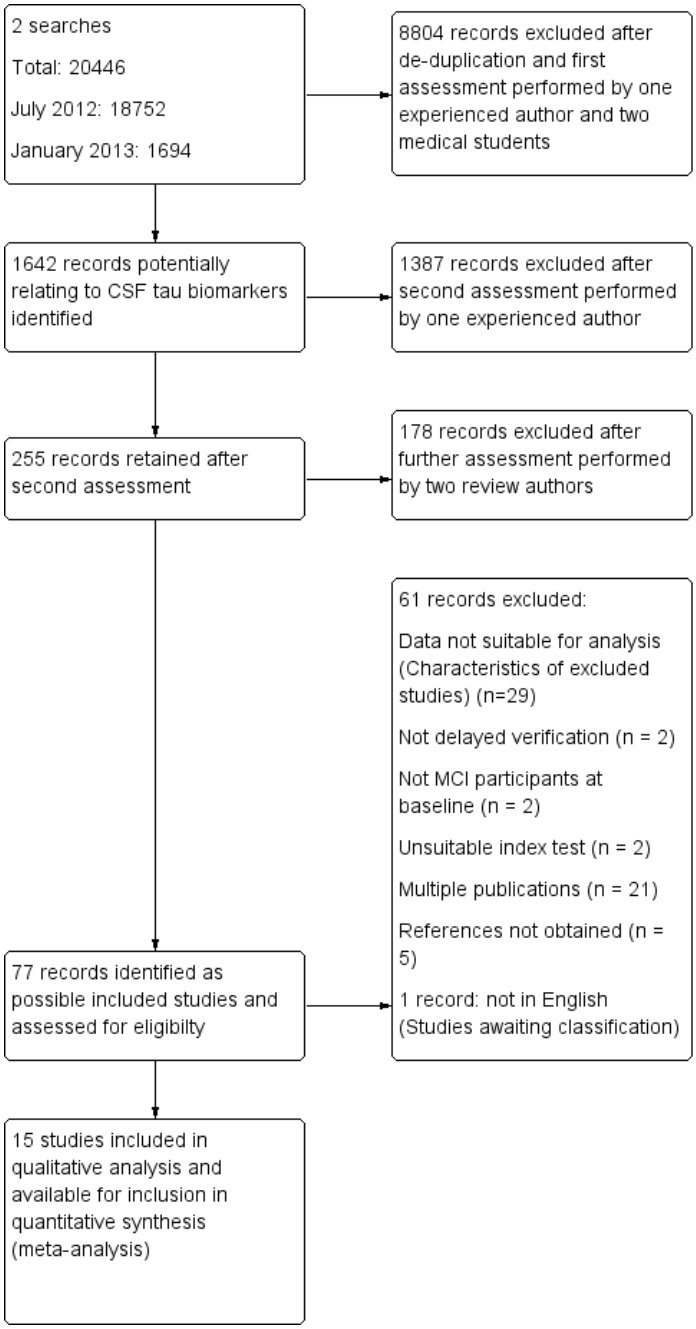
Two researchers (EL and AN-S) screened all titles and abstracts generated by the electronic database searches for relevance.

Two researchers (EL and AN-S) independently reviewed the remaining abstracts of selected titles and selected all potentially-eligible studies for full text review. Four researchers (NS, AN-S, SM and EL) independently further assessed full manuscripts against the inclusion criteria (see Criteria for considering studies for this review). Where necessary, a third arbitrator (CWR) resolved disagreements that the two researchers could not resolve through discussion.

Where a study included useable data but these were not presented in the published manuscript, we contacted the authors directly to request further information. If the same data set was presented in more than one paper, we included only the primary paper.

We detailed the numbers of studies selected at each point in a study flow diagram (Figure 1).

Figure 1. Study flow diagramNote: a top-up search performed in December 2015 revealed 6134 records85 records retained after de-duplication and assessment by one experienced reviewer81 records excluded after further assessment performed by two review authors4 studies identified for possible inclusion (Characteristics of studies awaiting classification)



Data extraction and management

We extracted data onto a study-specific form which included the following:

- Author, year of publication, and journal.
- The index test and assay type used (thresholds used to define positive and negative tests).
- The criteria used for clinical definition for the baseline population.
- Baseline demographics of the study population (age, gender, apolipoprotein E (ApoE) status, MMSE and clinical setting).
- The duration of follow-up (mean, minimum, maximum, and median).
- The proportion of participants developing the outcome of interest (Alzheimer's disease dementia using NINCDS-ADRDA criteria) as well as other forms of dementias where standard criteria were used.
- The sensitivity and specificity of the index test in defining Alzheimer's disease dementia (these were used to back-translate into a 2 x 2 table (Appendix 2)).
- Other data relevant for creating 2 x 2 tables (TP = true test positive; FP = false test positive; FN = false test negative; TN = true test negative) e.g. the number of 'abnormal' and 'normal' tests and baseline variables; the number of disease 'presence' and disease 'absence' at follow-up, as well as through scrutiny of scatter plots.

We also extracted data necessary for the assessment of quality as defined below.

Data extraction was performed independently by two blinded review authors (NS and AN-S). Disagreement in data extraction was resolved by discussion, with the potential to involve a third author (CWR) as arbitrator, if necessary.

Assessment of methodological quality

Two review authors (NS and AN-S), blinded to each other's scores, independently performed methodological quality assessments of each study using the QUADAS-2 tool (Whiting 2011), as recommended by the Cochrane Collaboration. Disagreement was resolved by further review and discussion with the potential to involve a third author (CWR) as arbitrator, if necessary.

The tool is made up of four domains: participant selection, index test, reference standard and participant flow. Each domain was assessed in terms of risk of bias, with the first three domains also considered in terms of applicability concerns (Quadas-2) (Appendix 3).The components of each of these domains and a rubric which details how judgments concerning risk of bias are made are detailed in Appendix 4. Certain key areas important for this review regarding quality assessment were participant selection, index test, and blinding.

We did not use QUADAS-2 data to form a summary quality score. We produced a narrative summary describing numbers of studies that were found to have high/low/unclear risk of bias, as well as concerns regarding applicability.

Statistical analysis and data synthesis

We evaluated test accuracy according to the target condition. There are no accepted thresholds to define what constitutes a positive or negative CSF index test for identifying those people with MCI who would convert to Alzheimer's disease dementia or other forms of dementia over time. Therefore, the estimates of diagnostic accuracy reported in primary studies were likely to be based on data-driven threshold selection (Leeflang 2008). We conducted exploratory analyses by plotting estimates of sensitivity and specificity from each study on forest plots and in receiver operating characteristic (ROC) space. We did not meta-analyse pairs of sensitivity and specificity using the bivariate model, as originally planned, because the results were not clinically interpretable when studies with mixed thresholds were included in the analysis. Instead, we fitted HSROC meta-analysis models to estimate summary ROC curves using SAS (Statistical Analysis Software), version 9.2 (SAS Institute 2011). We derived estimates of sensitivity and likelihood ratios at a fixed value of specificity (chosen a priori as the median specificity for the studies that were analysed when fitting the model) from the HSROC models for the illustrative purposes. Confidence intervals for sensitivity and the likelihood ratios were calculated using the delta method (Davison 2003), using the 'estimate' command after fitting the HSROC models in SAS. HSROC models were only fitted for analyses where data for 2 x 2 tables were provided by at least six studies, given the need to estimate five parameters. Where HSROC models were fitted, we summarised the post-test probability of conversion from MCI to dementia given a positive test result and given a negative test result for a range of prevalences of conversion (pretest) probabilities. This was done by plotting the post-test probabilities against the pretest probabilities, calculating the former based on the pretest probabilities and the likelihood ratios estimated from the HSROC model at the median of the observed specificity values from the included studies. A positive predictive value (PPV) and a negative predictive value (NPV) were also reported, based on the median prevalence (pretest probability) of conversion across studies. We caution that these post-test probabilities and PPV and NPV values related to likelihood ratios for hypothetical values of sensitivity and specificity for which the true threshold value of the index test was not known.

Investigations of heterogeneity

Heterogeneity was investigated through visual examination of forest plots of sensitivities and specificities and through visual examination of the ROC plot of the raw data. The main sources of heterogeneity were thought likely to be reference standards used, participant sampling, index test methodology and aspects of study quality (particularly inadequate blinding).

There were insufficient studies, therefore we did not perform meta-regression (by including each potential source of heterogeneity as a covariate in the HSROC model) as planned (Differences between protocol and review).

Sensitivity analyses

We planned to investigate the effect of quality items (such as pre-specifying threshold) on the accuracy of index tests by undertaking sensitivity analyses. Due to the limited number of studies, we did not perform any sensitivity analyses (Differences between protocol and review)

Assessment of reporting bias

We did not investigate reporting bias because of current uncertainty about how it operates in test accuracy studies and the interpretation of existing analytical tools such as funnel plots.

RESULTS

Results of the search

The total number of records identified by the searches up to January 2013 was 20,446. After de-duplication, a small team of assessors performed a first assessment of the remaining records. After a second assessment, 255 records were retained, of which 178 were excluded after assessment performed by two review authors. Seventy-seven references were identified as possible eligible studies and were assessed for inclusion (Figure 1). Fifteen papers were included, and 40 were discarded for the following reasons: i) data not suitable for analysis or insufficient data for creating two by two tables (n = 28) (Characteristics of excluded studies); ii) not a delayed verification study (n = 2); iii) not MCI participants at baseline (n = 2); iv) unsuitable index test (n = 2); v) reference not obtained (n = 5). In addition, twenty two papers were identified as multiple publications. One paper was not in English (Urakami 2004). No extra studies were found through reference checking. We obtained usable data for five papers (Amlien 2013; Galluzzi 2010; Hansson 2006; Visser 2009; Vos 2013) through contacting the authors.

We ran a top-up search in December 2015. The results of this search will be fully incorporated into the review at update. However, readers may wish to know that this search identified a total of 6314 results. After screening, four new studies were identified for inclusion within the review (please see Additional Tables: Table 1 for more details). The characteristics of these four new studies and their heterogeneity were all consistent with the fully incorporated studies.

Included Studies

The Characteristics of included studies table lists the characteristics of the 15 included studies containing a total of 1282 participants with MCI at baseline of whom 1172 had analysable data. Two studies (Buchhave 2012; Hansson 2006) involved the same cohort. Buchhave 2012 reported the data for the CSF p-tau/ABeta ratio index test from a new follow-up period.

Study designs were seven prospectively well-defined cohorts of participants with MCI (Buchhave 2012; Fellgiebel 2007; Galluzzi 2010; Herukka 2007; Kester 2011; Palmqvist 2012; Vos 2013), six nested case-control studies with a prospectively defined MCI group (Amlien 2013; Hansson 2006; Koivunen 2008; Monge-Argiles 2011; Parnetti 2012; Visser 2009) and two studies with a retrospectively defined MCI group with longitudinal data (Eckerstrom 2010; Hampel 2004).

A majority of studies (n = 9) were published between 2010 and 2013. The remaining six studies were published from 2004 to 2008. All of the included studies were conducted in Europe (five in Sweden, two in Italy and two in Finland, one in The Netherland, one in Spain, one in Norway, one in Germany and two were European multi-centre studies). They used one version or another of the Petersen criteria for MCI. Twelve studies applied NINCDS-ADRDA criteria or NINCDS-ADRDA and DSM criteria as a reference standard for Alzheimer's disease dementia. Amlien 2013 used Global Dementia Scale (GDS) & Research criteria, Fellgiebel 2007 used 'CDR = 1 criteria' and Parnetti 2012 did not specify the reference standard at follow-up.

Study sizes varied and ranged from 15 (Koivunen 2008) to 231 participants (Vos 2013). Nine papers had included participants with a mean age of 70 years or under. The mean (range) age of the youngest sample was 64 years (45 to 76) (Amlien 2013) and the mean (SD) age of the oldest sample was 73.4 (6.6) years (Monge-Argiles 2011). Sampling procedure and APOE ε 4 gene carriers were poorly reported. Participants were mainly recruited from university memory clinics (n = 8), while one study did not report sources of recruitment (Koivunen 2008).

Different CSF biomarker level values were used as a threshold in the included studies (Additional tables: Table 2). The threshold was prespecified in only five studies (Amlien 2013; Herukka 2007; Kester 2011; Koivunen 2008; Vos 2013). A percentage of converters to Alzheimer's disease dementia ranged from 22% (Visser 2009) to 56% (Hampel 2004). CSF index test positivity ranged from 23% (Amlien 2013) to 69% (Vos 2013) . Duration of follow-up was reported as mean and standard deviation (SD), or median, or range. Most studies had follow-up between 12 and 36 months. Some participants were followed up for less than one

year in three of the included studies (Fellgiebel 2007; Hampel 2004; Monge-Argiles 2011), and for more than four years in five of the included studies (Buchhave 2012; Herukka 2007; Palmqvist 2012; Parnetti 2012). Participants in the remaining seven studies (Amlien 2013; Eckerstrom 2010; Galluzzi 2010; Kester 2011; Koivunen 2008; Visser 2009) were followed up from one to three years.

Excluded studies

Twenty-nine studies, nine of which were ADNI studies, were excluded as they failed to meet the inclusion criteria for participants, index test, target condition, or they didn't have diagnostic accuracy data (Characteristics of excluded studies). We contacted the authors of two of the ADNI studies (Landau 2010; Westman 2012) in order to obtain additional data for creating two by two tables. Further information was not available for the Landau 2010 study at the time this review was prepared. The author of the Westman 2012 study informed us that the accuracy of combined, not individual, CSF biomarkers was assessed in their study.

Studies awaiting classifications

The Characteristics of studies awaiting classification table lists the characteristics of four studies which might be considered for the inclusion in an updated review. The authors of all those studies need to be contacted in order to obtain missing data/relevant information. Regarding the target condition 'Conversion from MCI to Alzheimer's disease', provisional data from two studies (Ewers 2012; Leuzy 2015) might be used for the analysis of CSF t-tau; data for the analysis of CSF p-tau ABeta42/p-tau ratio index tests might be available only from Ewers 2012 and Balasa 2014, respectively.

Additional Tables: Table 1 shows that the percentage of converters to Alzheimer's disease dementia ranged from 36% to 47%. Duration of follow-up was between 24 and 41 months. Leuzy 2015 did not report duration of follow-up and Ewers 2012 did not report a threshold value. The heterogeneity of results in these four studies was consistent with that observed in the fully incorporated studies.

Methodological quality of included studies

Methodological quality was assessed using the QUADAS-2 tool (Whiting 2011).

Review authors' judgements about each methodological quality item for each included study are presented in the Characteristics of included studies table and Figure 2. The overall methodological quality of included study cohorts is summarised in Figure 3.

Figure 2. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study

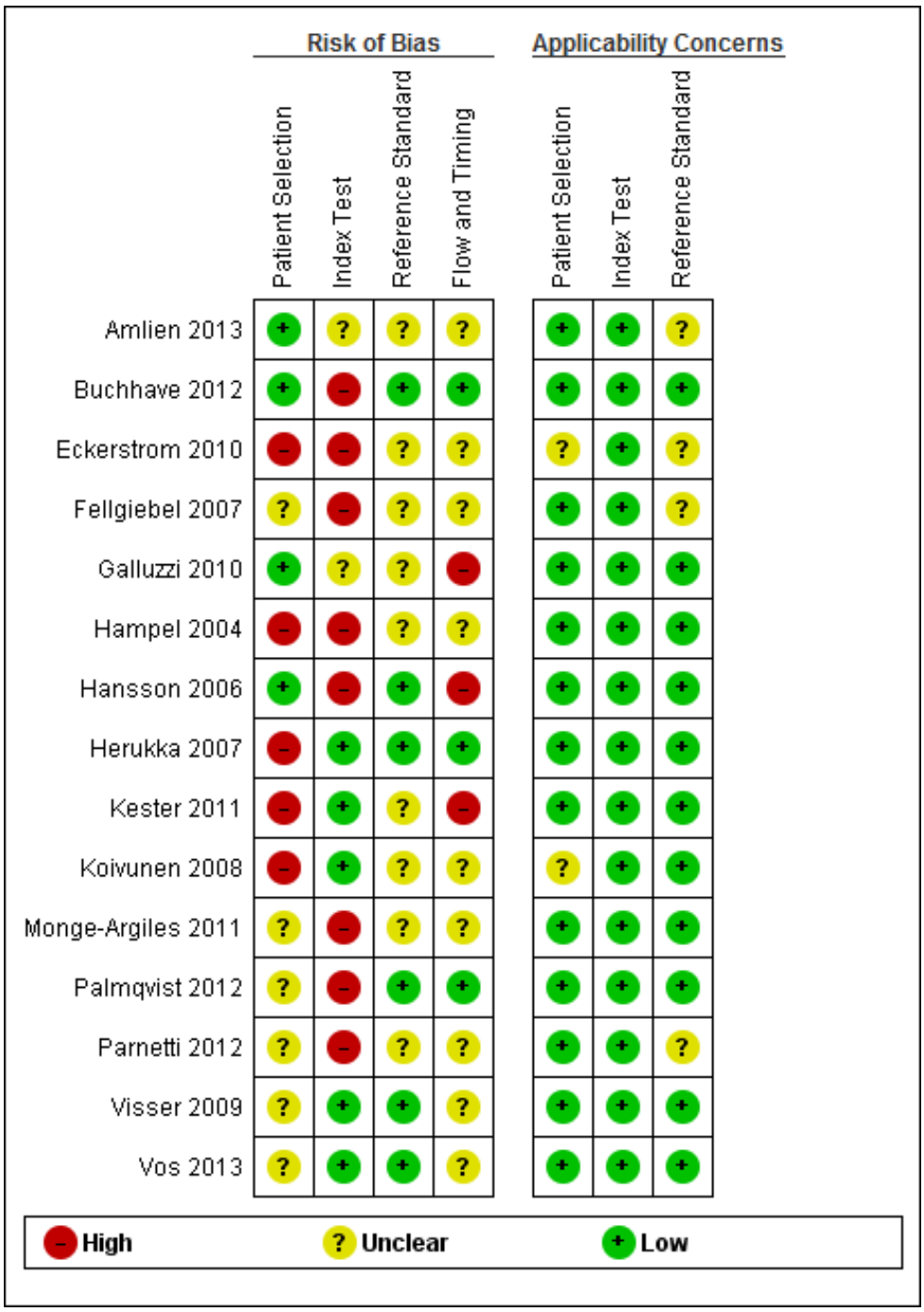
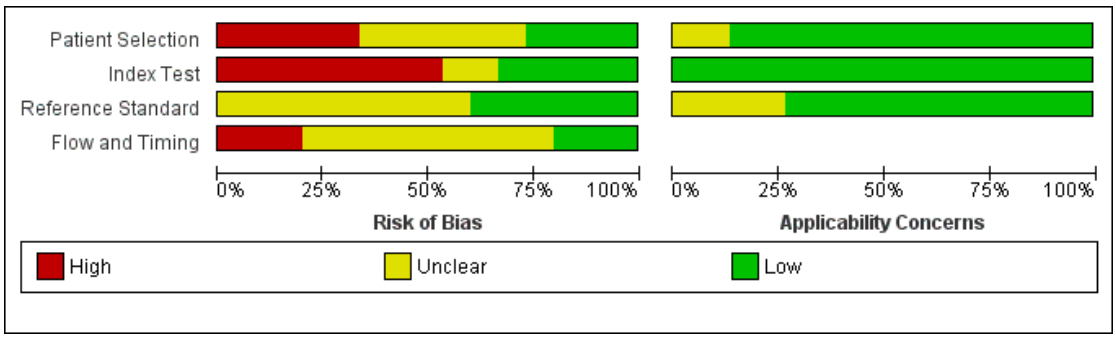


Figure 3. Risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across included studies



In the participant selection domain, we considered five studies (Eckerstrom 2010; Hampel 2004; Herukka 2007; Kester 2011; Koivunen 2008) to be at high risk of bias because the participants were not consecutively or randomly enrolled or both the sampling procedure and exclusion criteria were not described. We stated that all included studies avoided a case-control design because we only considered data on performance of the index test to discriminate between people with MCI who converted to dementia and those who remained stable. We considered four studies (Amlien 2013; Buchhave 2012; Galluzzi 2010; Hansson 2006) to be at low risk of bias. We considered the remaining six studies to be at unclear risk of bias, due to poor reporting on sampling procedure or exclusion criteria

In the index test domain, we considered eight studies (Buchhave 2012; Eckerstrom 2010; Fellgiebel 2007; Hampel 2004; Hansson 2006; Monge-Argiles 2011; Palmqvist 2012; Parnetti 2012) to be at high risk of bias because the threshold used was not prespecified and the optimal cutoff level was determined from ROC analyses; therefore, the accuracy of the CSF biomarkers reported in these studies appeared to be overestimated. We considered two studies (Amlien 2013; Galluzzi 2010) to be at unclear risk of bias, due to poor reporting. We considered the remaining five studies to be at low risk of bias.

In the reference standard domain, we considered nine studies (Amlien 2013; Eckerstrom 2010; Fellgiebel 2007; Galluzzi 2010; Hampel 2004; Kester 2011; Koivunen 2008; Monge-Argiles 2011; Parnetti 2012) to be at unclear risk of bias, mainly because it was not reported whether clinicians conducting follow-up were aware of initial CSF biomarker analysis results. Three of those nine studies did not clearly report the reference standards used for diagnosing Alzheimer's disease dementia. We were not able to obtain the information about how the reference standard was obtained and by whom, due to poor reporting. We considered the remaining six studies to be at low risk of bias.

In the flow and timing domain, we judged nine studies (Amlien 2013; Eckerstrom 2010; Fellgiebel 2007; Galluzzi 2010; Koivunen 2008; Monge-Argiles 2011; Parnetti 2012; Visser 2009; Vos 2013) to be at unclear risk of bias because not all participants were included in the analysis and/or the follow-up period was shorter than one year and/or reporting was poor. We judged three studies (Galluzzi 2010; Hansson 2006; Kester 2011) to be at high risk of bias because a large number of participants with non-Alzheimer's disease dementia were excluded from the analysis. We considered the remaining three studies to be at low risk of bias.

For assessment of applicability concerns, for the majority of the studies there was no concern that the included participants and setting, the conduct and interpretation of the index test, and the target condition (as defined by the reference standard) in each of the included studies did not match the review question. We judged two studies (Eckerstrom 2010; Koivunen 2008) to be of unclear applicability because of concerns regarding the participant characteristics or setting. We also judged four studies (Amlien 2013; Eckerstrom 2010; Fellgiebel 2007; Parnetti 2012) to be of unclear applicability because of concerns with respect to the reference standard.

It should be noted that the lack of concern about applicability of the three domains mentioned above was based on the inclusion criteria set in the review, and therefore the judgment about applicability may be overstated.

Findings

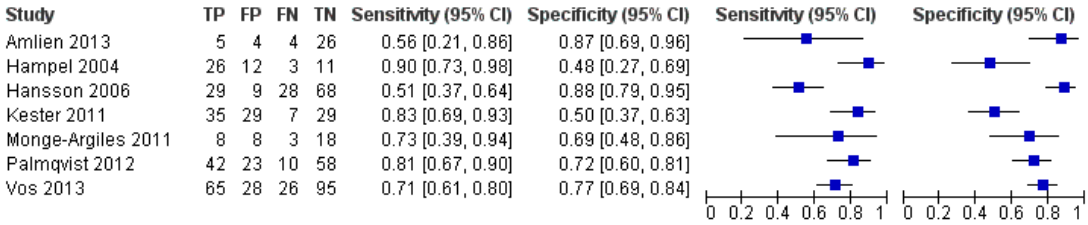
The key characteristics of each study are summarised in Additional

Tables: Table 2 and Table 3. Included studies used a range of different thresholds. The number of positive CSF index tests at baseline varied across studies. The summary of main results for the fifteen included studies is presented in Summary of findings.

CSF t-tau for Alzheimer's disease dementia

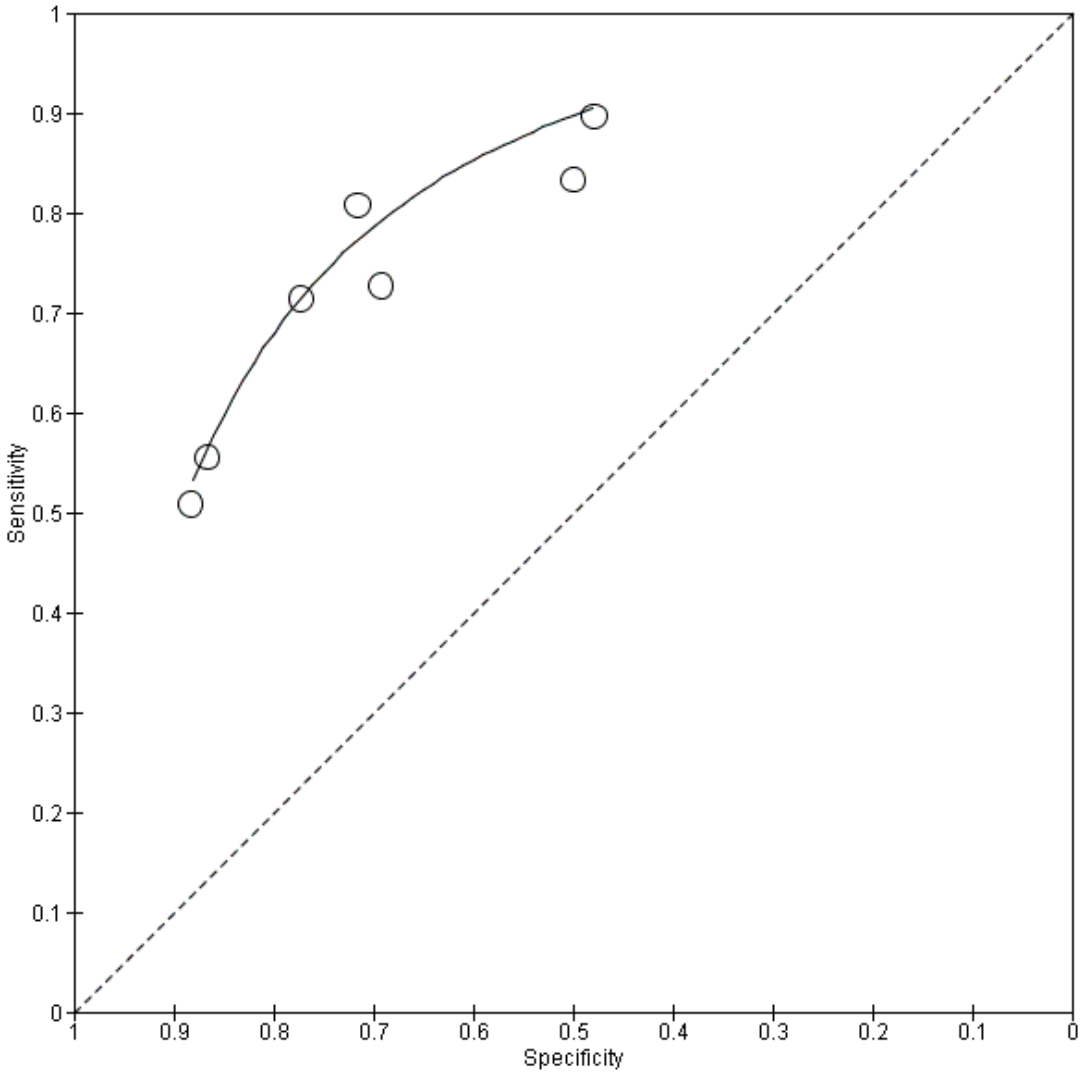
Individual study estimates of sensitivity and specificity are shown in Figure 4 for each of the seven studies (291 cases and 418 non-cases) that evaluated Alzheimer's disease dementia. The sensitivity values ranged from 51% to 90% while the specificity values ranged from 48% to 88%. The thresholds used ranged from ≥ 77 to ≥ 500 pg/mL (ng/L).

Figure 4. Forest plot of 1 CSF t-tau conversion to AD dementia.



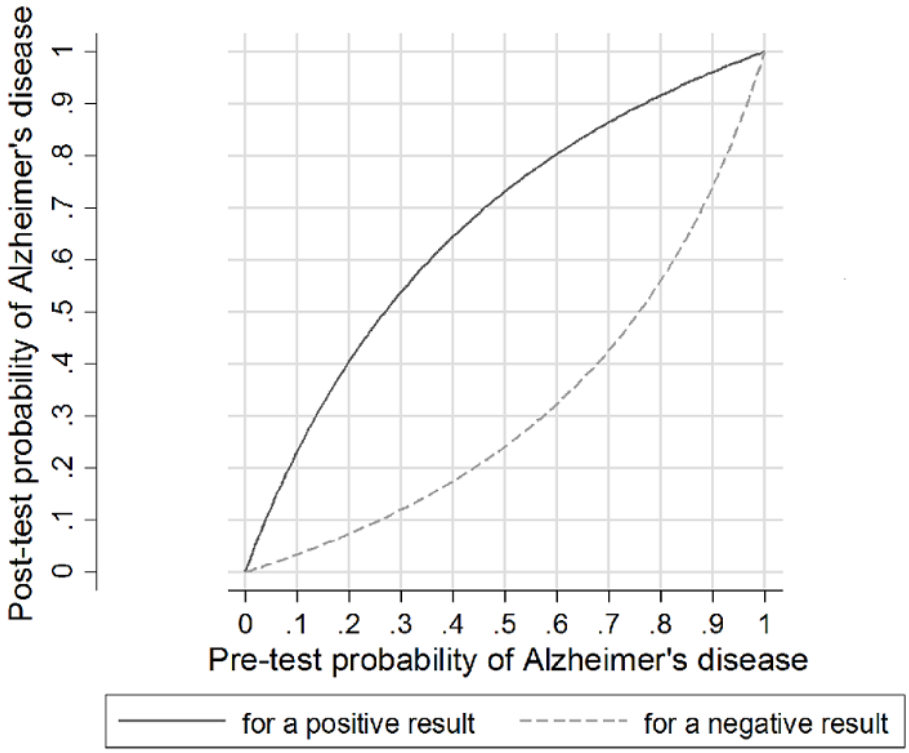
The summary ROC curve summarising the accuracy of CSF t-tau across the seven studies is shown in Figure 5. Because of the variation in thresholds, we did not estimate a summary sensitivity and specificity. However, we derived estimates of sensitivity and likelihood ratios at fixed values of specificity from the HSROC model we fitted to produce the summary ROC curve. At the median specificity of 72%, the estimated sensitivity was 77% (95% CI 67 to 85), the positive likelihood ratio was 2.72 (95% CI 2.43 to 3.04), and the negative likelihood ratio was 0.32 (95% CI 0.22 to 0.47).

Figure 5. Summary ROC Plot of 1 CSF t-tau conversion to AD dementia.



At the median specificity (72%) and the median prevalence of Alzheimer's disease dementia (37%) (pretest probability, Figure 6), the positive predictive value was 62%, which means on average 62 out of 100 people with MCI and a positive index test result would convert to Alzheimer's disease dementia, but 38 would not. The negative predictive value of 84% means that on average 84 out of 100 people with MCI and with a negative index test result would not convert to Alzheimer's disease dementia, but 16 would.

Figure 6. Post-test probability plots (Analysis 1): Conversion from MCI to Alzheimer's disease for CSF t-tau as a diagnostic test



In a hypothetical cohort of 100 people with MCI taking the CSF t-tau test, there would be on average nine false negatives (participants who convert but incorrectly tested negative) and 18 false positives (participants who did not convert but incorrectly tested positive) (Summary of findings).

CSF p-tau for Alzheimer's disease dementia

Six studies (164 cases and 328 non-cases) evaluated the accuracy of CSF p-tau for conversion to Alzheimer's disease dementia (Figure 7). The sensitivities were between 40% and 100%, while the specificities were between 22% and 86%. The thresholds used ranged from ≥ 39 to ≥ 85 pg/mL (ng/L).

Figure 7. Forest plot of 2 CSF p-tau conversion to AD dementia.

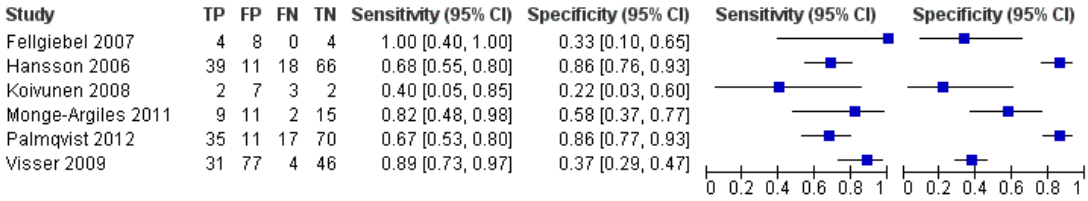
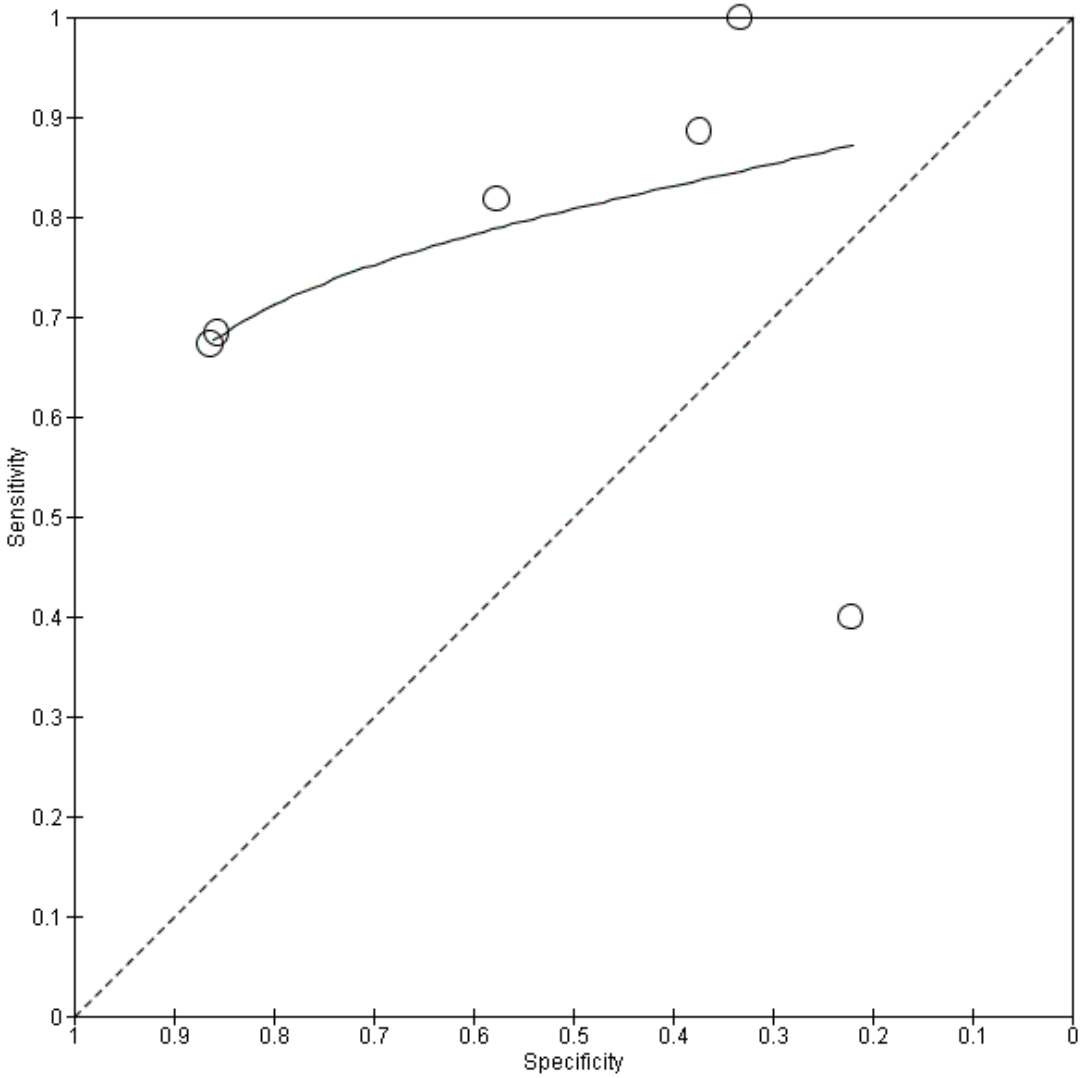


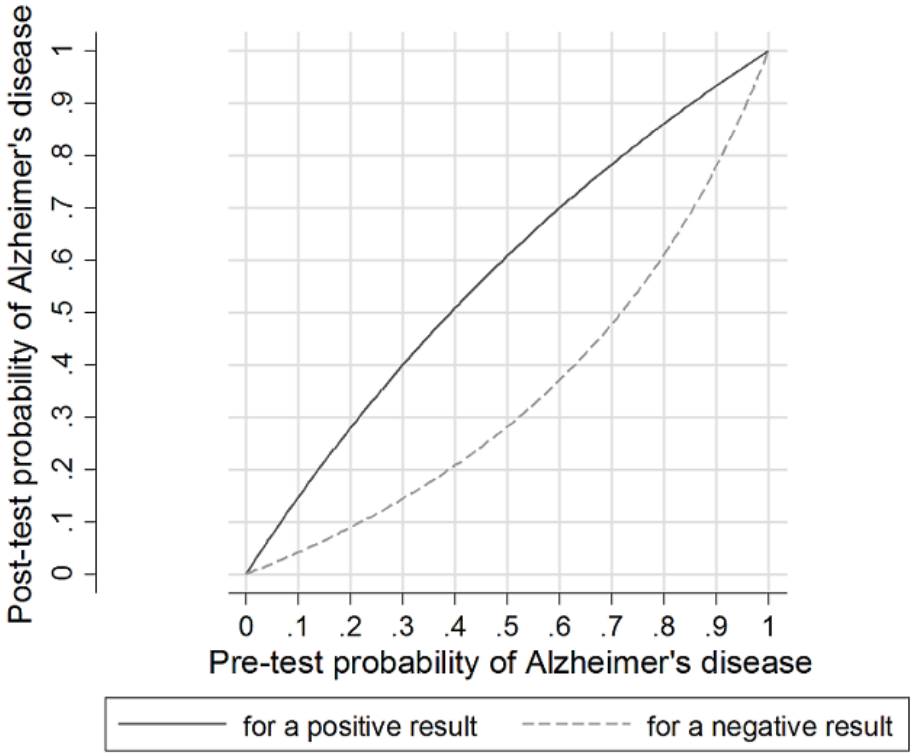
Figure 8 shows the summary ROC space. We derived the summary estimates at different points on the fitted HSROC curve. At the median specificity of 48%, the estimated sensitivity was 81% (95% CI 64 to 91), the positive likelihood ratio was 1.55 (CI 1.31 to 1.84), and the negative likelihood ratio was 0.39 (CI 0.19 to 0.82).

Figure 8. Summary ROC Plot of 2 CSF p-tau conversion to AD dementia.



At the median specificity (48%) and the median prevalence of Alzheimer's disease dementia (37%) (pretest probability, Figure 9), the positive predictive value was 48%, which means on average 48 out of 100 people with MCI and with a positive index test result would convert to Alzheimer's disease dementia but 52 would not. The negative predictive value of 81% means that on average 81 out of 100 people with MCI with a negative index test result would not convert to Alzheimer's disease dementia, but 19 would.

Figure 9. Post-test probability plots (Analysis 2): Conversion from MCI to Alzheimer's disease for CSF p-tau as a diagnostic test



In a hypothetical cohort of 100 people with MCI taking the CSF p-tau test, there would be on average seven false negatives (participants who convert but incorrectly tested negative) and 33 false positives (participants who did not convert but incorrectly tested positive) (Summary of findings).

CSF p-tau/ABeta ratio for Alzheimer's disease dementia

Five studies (140 cases and 293 non-cases) evaluated the accuracy of the CSF p-tau/ABeta ratio for conversion to Alzheimer's disease dementia (Figure 10). The sensitivities were between 80% and 96%, while the specificities were between 33% and 95%. We were not able to report the range of thresholds due to different measurements: < 6.6 pg/mL (ng/L); 0.18; 1074.0; < 9.92. Figure 11 shows the summary ROC space. We did not conduct a meta-analysis because the studies were few and small.

Figure 10. Forest plot of 3 CSF p-tau/ABeta ratio to AD dementia.

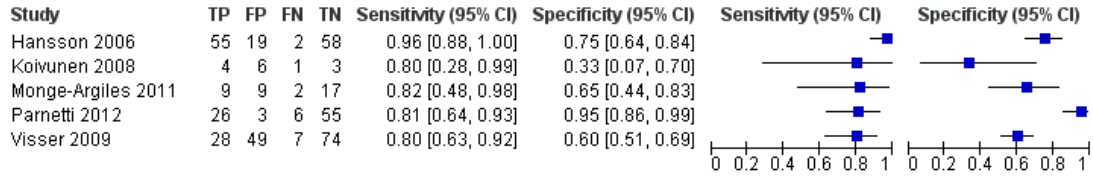
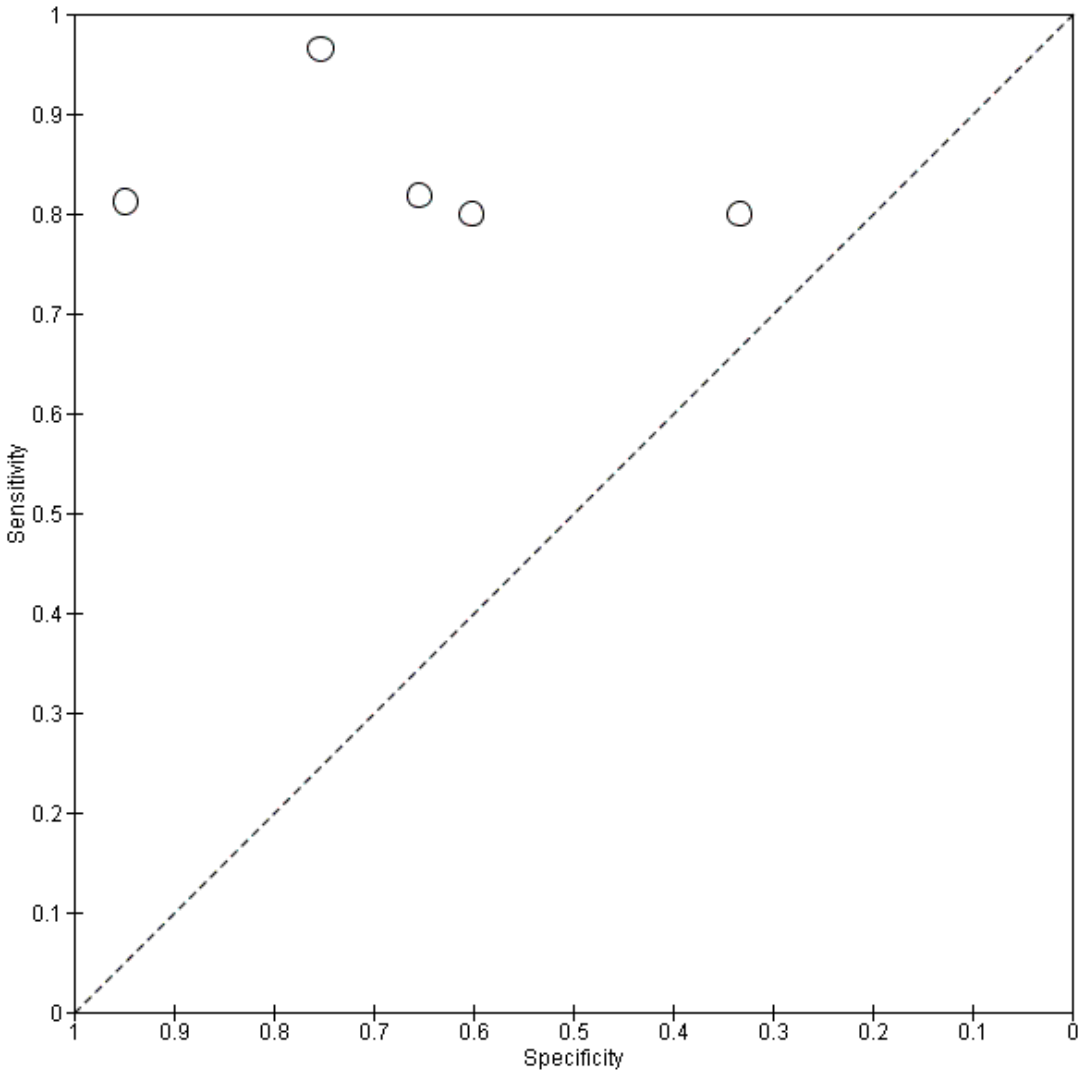


Figure 11. Summary ROC Plot of 3 CSF p-tau/ABeta ratio to AD dementia.



CSF t-tau/ABeta ratio for Alzheimer's disease dementia

Only two studies (Monge-Argiles 2011; Vos 2013) evaluated the accuracy of the CSF t-tau/ABeta ratio for conversion to Alzheimer's disease dementia. The sensitivities were 50% and 51%, and specificities were 91% and 96%, respectively. We were not able to conduct the meta-analysis.

CSF t-tau for all forms of dementia (combined Alzheimer's disease dementia and non-Alzheimer's disease dementia)

Only four studies (166 cases and 153 non-cases) evaluated the accuracy of CSF t-tau for conversion to all forms of dementia (Figure 12 and Figure 13). The sensitivity values ranged from 42% to 79%, while the specificity values ranged from 63% to 95%.

The thresholds used ranged from > 350 to ≥ 500 pg/mL (ng/L). As above, we did not conduct a meta-analysis because the studies were few and small.

Figure 12. Forest plot of 4 CSF t-tau conversion to all forms of dementia.

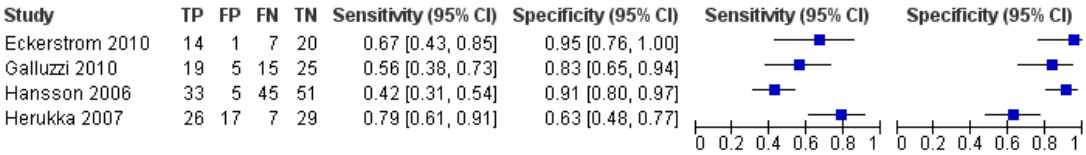
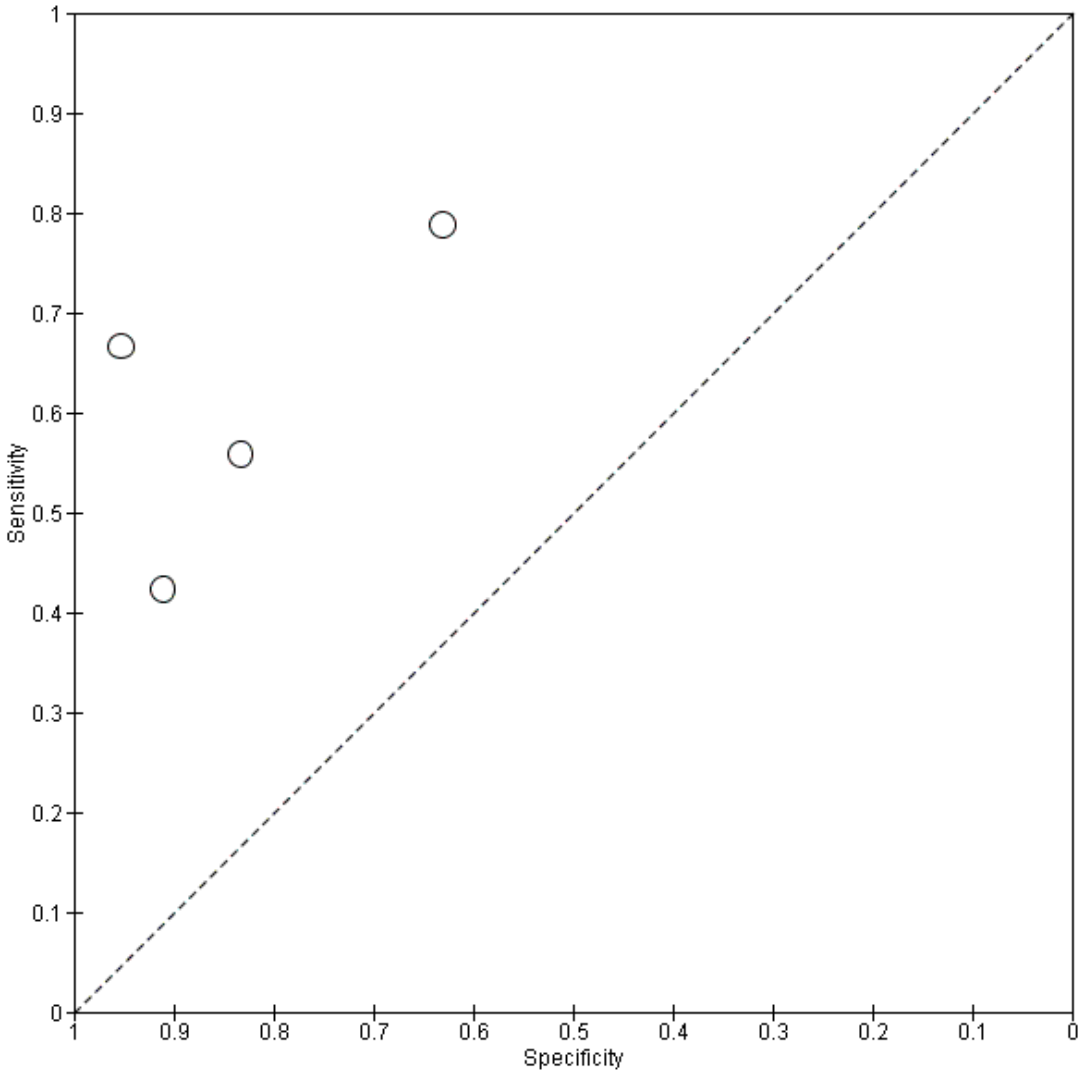


Figure 13. Summary ROC Plot of 4 CSF t-tau conversion to All dementias.



Investigation of heterogeneity

We were not able to formally assess the effects of each potential source of heterogeneity as planned, due to the small number of studies available to be included.

Sensitivity analyses

Due to the limited number of studies evaluating each of four CSF biomarkers for Alzheimer's disease dementia or other types of dementia, we did not perform any sensitivity analyses, as planned.

Summary of findings

What is the diagnostic accuracy of CSF biomarker levels for detecting Alzheimer's disease pathology in people with mild cognitive impairment (MCI), and identifying those MCI participants who would convert to Alzheimer's disease dementia or other forms of dementia over time	
Descriptive	
Patient population	Participants diagnosed with MCI at baseline using any of the Petersen criteria or CDR = 0.5 or any 16 definitions included by Matthews (Matthews 2008)
Sampling procedure	Consecutive or random (n = 5) Not consecutive or random (n = 3) Unclear (n = 7)
Sources of recruitment	University memory clinic (n = 8); European multicentre memory clinics (n = 2); inpatients (n = 2); General Hospital memory clinic (n = 1); Research centre outpatient memory clinic (n = 1); not reported (n = 1)
Prior testing	The only testing prior to performing the plasma and CSF biomarkers was the application of diagnostic criteria for identifying participants with MCI
MCI criteria	Petersen criteria (n = 14) Global Deterioration Scale (GDS) (n = 1)
Index tests	CSF t-tau or CSF p-tau or CSF p-tau/ABeta ratio or CSF t-tau/ABeta ratio
Reference standard	NINCDS-ADRDA and/or DSM and/or ICD criteria for Alzheimer's disease dementia (n = 12); Global Dementia Scale (GDS) & Research criteria (n = 1); CDR = 1 criteria (n = 1); not specified (n = 1) McKeith criteria for Lewy body dementia; Lund criteria for frontotemporal dementia; and NINDS AIREN criteria for vascular dementia
Target condition	Alzheimer's disease dementia or any other types of dementia
Included studies	Prospectively well-defined cohorts of MCI participants (n = 7), nested case-control studies with a prospectively defined MCI group (n = 6) and studies with a retrospectively defined MCI group with longitudinal data (n = 2) Fifteen studies (N = 1282 participants) were included. Number included in analysis: 1172
Quality concerns	Patient selection and conduct of the reference standard were poorly reported. Applicability concerns were generally low. Regarding the inclusion criteria set in the review, the majority of included studies did match the review question*: 'Could CSF t-tau and CSF t-tau/ABetaratio biomarkers identify those MCI participants with Alzheimer's disease pathology at baseline who would convert clinically to dementia at follow up?'* However, due to a limited number of included studies and levels of heterogeneity, it is difficult to determine to what extent the findings from a meta-analysis can be applied to clinical practice

Limitations	Limited investigation of heterogeneity due to insufficient number of studies. There was a lack of common thresholds						
Test Median percentage converting (range) [2]	**Studies**	**Cases/participants**	**Median specificity from included studies**	**Sensitivity (95% CI)**[1] **at median specificity**	**Consequences in a cohort of 100**		
					Median percentage converting[2]	**Missed cases**	**Overdiagnosed**
Alzheimer's disease dementia							
CSF t-tau	7	436/709	72	77 (67, 85)	37	9	18
Alzheimer's disease dementia							
CSF p-tau	6	164/492	47.5	81 (64, 91.5)	37	7	33
Alzheimer's disease dementia							
CSF p-tau/ ABeta ratio	5	140/433	No meta-analysis	No meta-analysis			
All types of dementia							
CSF t-tau	4	166/319	No meta-analysis	No meta-analysis			
Investigation of heterogeneity: the planned investigations were not possible due to the limited number of studies available for each analysis. We were unable to investigate the effect of duration of follow-up due to substantial variation in length and reporting							
Conclusions: Given the insufficient evidence to evaluate the diagnostic value in MCI of CSF t-tau, CSF p-tau, CSF t-tau/ABeta ratio and CSF p-tau/ABeta ratio for Alzheimer's disease dementia and other forms of dementias examined in this review, particular attention should be paid to the risk of misdiagnosis and overdiagnosis of dementia (and therefore overtreatment) in clinical practice. Future studies with more uniform approaches to thresholds, analysis and study conduct may provide a more homogenous estimate than the one that has been available from the included studies we have identified							

[1]Meta-analytic estimate of sensitivity derived from the HSROC model at a fixed value of specificity. Summary estimates of sensitivity and specificity were not computed because the studies that contributed to the estimation of the summary ROC curve used different thresholds.

[2]The median percentage converting was calculated using all the studies that reported 'conversion from MCI to Alzheimers' disease dementia' (Table 2)

DISCUSSION

We performed a review of the available evidence on the diagnostic accuracy of CSF biomarker levels for detecting Alzheimer's disease pathology in people with MCI, and identifying those MCI participants who would convert to Alzheimer's disease dementia or other forms of dementia over time. In the absence of a contemporaneous reference standard for Alzheimer's disease diagnosis relative to the application of the index test, the decision to use a delayed verification design was taken for all DTA reviews by our group. This, however, creates problems when the length of follow-up in studies varies, as the longer a study, in a chronic disorder where age is the principal risk factor, could create false positive findings. To address this, length of follow-up was collected to help interpret between-study variations in accuracy.

There is, however, a paucity of evidence in relation to the accuracy of CSF biomarkers. Where data were available for conversion to Alzheimer's disease dementia, there was a wide range of sensitivity (51% to 90%; 40% to 100%; 80% to 96%) and specificity (48% to 88%; 22% to 86%; 33% to 95%) values for the CSF t-tau, CSF p-tau and CSF p-tau/ABeta ratio index tests, respectively.

Due to the wide variations in thresholds, we did not estimate a summary sensitivity and specificity. Although, subject to considerable uncertainty of a statistical approach, in order to illustrate the potential strengths and weaknesses of CSF biomarker levels we estimated from the fitted summary ROC curve that the sensitivity was 77% (95% CI 67 to 85) and 81% (95% CI 64 to 91) at the included study median specificity of 72% and 48% for the CSF t-tau and CSF p-tau respectively. Assuming a conversion rate of MCI to Alzheimer's dementia of 37%, for every 100 CSF t-tau level, nine individuals with a negative test would progress and 18 with a positive test would not progress to Alzheimer's dementia; for every 100 CSF p-tau level, seven individuals with a negative test would progress and 33 with a positive test would not progress to Alzheimer's dementia. The estimation of predictive values and consequences in a cohort of 100 ('missed cases' and 'over-diagnosed') were based on hypothetical sensitivity and specificity values for which the threshold of the test is unknown; therefore, these findings should be interpreted with caution.

We were not able to evaluate the accuracy of CSF biomarkers for conversion from MCI to other forms of dementia (non-Alzheimer's disease dementia). As a result of the information available from four studies (Eckerstrom 2010; Galluzzi 2010; Hansson 2006; Herukka 2007), we evaluated the accuracy of CSF t-tau for conversion to all types of dementia (combined Alzheimer's disease dementia and non-Alzheimer's disease dementia). The sensitivity values ranged from 42% to 79% while the specificity values ranged from 63% to 95%. We did not conduct a meta-analysis because the studies were few and small.

Previous reviews of tests of amyloid in CSF and plasma (Ritchie 2014) and evidenced through PET imaging (Zhang 2014) have been published. They highlighted that as a test, there was consistently better sensitivity than specificity whereby the absence of evidence of amyloid pathology (low levels in CSF and high levels in the cortices) was likely to exclude a later diagnosis of Alzheimer's disease dementia, whereas the presence of amyloid pathology did not add much incremental benefit to diagnostic accuracy. Considering the findings of this systematic review, we have demonstrated again that the NPV is greater than the PPV which is a reflection of the higher sensitivity of these tests compared to their specificity. That is, a test indicating absence of biomarker abnormality and hence suggesting absence of disease is of more value than a positive biomarker indicating disease. CSF biomarkers are better at ruling out Alzheimer's disease than ruling it in as a cause of the clinical symptoms, and therein progression to Alzheimer's dementia in people described as having MCI. However, the reported optimal thresholds in individual papers tended to yield better sensitivities than specificities and this was reflected in our sROC analysis; therefore, those results should be interpreted with caution.

Given the insufficient evidence to evaluate the diagnostic value in MCI of CSF t-tau, CSF p-tau and the CSF p-tau/ABeta ratio for Alzheimer's disease dementia and other dementias examined in this review, particular attention should be paid to the risk of misdiagnosis and overdiagnosis of dementia (and therefore overtreatment) in clinical practice. Our findings are consistent with the expert opinion conveyed by Molinuevo et al (Molinuevo 2014) where it was recognised that negative tests results were more clinically useful than positive ones. They still saw a routine use for these tests in clinical practice, and our review will help describe the degree of accuracy to help inform clinicians using this test in their current practice. As sensitivity of this test was better than specificity, the risk of a missed diagnosis, or a false-negative test was lower. False reassurance given to a patient that they don't have or will not get Alzheimer's dementia would also have serious clinical consequences; however, appropriate pretest counselling for what can and cannot be revealed through CSF testing would mitigate the risk of an inappropriate level of salience being afforded to this particular test.

Summary of main results

In total, 1282 participants with MCI at baseline were identified in the fifteen included studies, of which 1172 had analysable data; 430 participants converted to Alzheimer's disease dementia and 130 participants to other forms of dementia at follow-up. It was possible to undertake a summary analysis of the CSF t-tau and p-tau markers but not the ratio, as too few studies presented results for the ratio. Consistent with the findings from the amyloid reviews, CSF t-tau and p-tau were reasonably sensitive tests for later diagnosis of Alzheimer's disease dementia, but had poor specificity. This is illustrated in Figure 6 and Figure 9 where the

small positive likelihood ratio for both CSF t-tau and p-tau has very little impact on the change from pretest probability to post-test probability. With respect to the CSF t-tau/Abeta ratio, it was not possible to generate likelihood ratios, due to only one study (Monge-Argiles 2011) reporting data . However, from Figure 10, it can be seen that for all but one study (Parnetti 2012), the sensitivity exceeded the specificity for the p-tau/ABeta ratio. Figure 12 though demonstrates across four studies, that the specificity of CSF t-tau is improved when the outcome is 'all forms of dementia', suggesting that the elevation of tau is a nonspecific marker of neurodegeneration and not tightly tethered to Alzheimer's disease pathology.

Our findings were based on studies with poor reporting and most included studies had an unclear risk of bias, mainly for reference standard and participant selection domains. Nine studies (56%) had unclear risk of bias for the flow and timing domain, mainly due to not including all participants in the analysis or inappropriate duration of the follow-up period. According to the assessment of the index test domain, 50% of studies were of poor methodological quality.

The main sources of heterogeneity were thought likely to be index test thresholds, reference standards used for the target disorders, sources of recruitment, participant sampling and aspects of study quality (particularly inadequate blinding). We were not able to formally assess the effects of each potential source of heterogeneity, as planned, due to the small number of studies available to be included.

Strengths and weaknesses of the review

There were a number of strengths to this review. This review was conducted in adherence to the inclusion criteria and methods described in a published protocol (Ritchie 2011). We searched a number of electronic databases, using an extensive range of appropriate database indexing terms and equivalent text words covering the index test, how it was measured, and the target condition. The multi-stranded search approach that we adopted to combine different search concepts in searches run in parallel, some including a more specific diagnostic component, has successfully increased the overall sensitivity of the search and is a strength of this review. Our searches were not limited by language. We contacted 12 study authors and usable data were obtained for five studies (Amlien 2013; Galluzzi 2010; Hansson 2006; Visser 2009; Vos 2013).

There were, however, also a number of limitations to this review. There was limited published information and substantial variation in the quality of the papers and caution is needed when interpreting these findings. Most included studies provided little data on participants at baseline. Several studies reported high or unclear dropout and withdrawal rates. Studies also contained wide variations in thresholds. It is also a weakness of the review that variability in length of follow-up in the various cohorts was so great. It would stand to reason that a longer follow-up period would more likely yield more cases of dementia, given that age is the principal risk factor for dementia. On the other hand, short follow-up periods might increase false negative results. This topic is of great interest to the field where determination of proximal and distal biomarkers are being considered. In an MCI population presenting to a clinician, it is the question of proximity to a decline to dementia which is the most relevant; in this regard, follow-up periods of over five years lose clinical meaningfulness. Standardisation of the follow-up period would help reviews like this; this has been suggested in our group's recent STARDdem proposals (Noel-Storr 2014). In our review, we were unable to formally test what affect length of follow-up had on the accuracy of the test. The various contributors to the heterogeneity across the studies may affect the study results. Given the poor reporting within the included studies, it is difficult to determine the underlying difference or differences among the included studies. This highlights a shortfall of large-scale, high-quality empirical research conducted in this area. Future studies should provide clearer reporting of the participants, equipment, usage and the implications of implementing the tests. As the current research area is rapidly changing, further research exploring the impact of the CSF p-tau/ABeta ratio on clinical outcomes is needed. To this end, we conducted a very recent literature review which revealed four new studies that will be fully incorporated in our next planned update. These four studies demonstrated the same between-study heterogeneity in results and methodology that we had observed in the included studies, with the implication that there will not be an impact of incorporation on our existing conclusions.

Applicability of findings to the review question

These findings can be considered a reasonable answer to the question being set in this review. Caution, though, should still apply because of the quality and reporting issues highlighted from the included papers and the small data set. This is especially true when drawing conclusions from the analysis of the p-tau/ABeta ratio. This is particularly important as it this ratio that is often favoured in clinical practice as being most accurate. However, this review and the previous published reviews of amyloid tests and Alzheimer's disease pathology consistently demonstrate reasonable sensitivity and poor specificity; accordingly, it is likely that the ratio of two sensitive tests will generate greater sensitivity than specificity.

AUTHORS' CONCLUSIONS

Implications for practice

The principal conclusion from our review is that of ongoing uncertainty regarding the true value of these tests in the management of people with prodromal dementia or MCI.

The use of and access to lumbar punctures (LP) in dementia clinics varies greatly between and within countries. The test is usually straightforward with only very occasional side effects, such as headache. However, acceptability of an LP by patients and carers also varies greatly and may reflect the views of the clinicians proposing their use and their perspective on the value of the test diagnostically. In the context of the new diagnostic criteria being used for prodromal Alzheimer's dementia (Dubois 2014), the tests studied here have been used as being indicative of Alzheimer's pathology. The data from this review suggests that a negative CSF test, in people with MCI, is likely to reflect the absence of Alzheimer's disease pathology as the aetiology of their clinical symptoms. However, in CSF sampling for ABeta and tau levels, a positive result does very little to indicate the presence of Alzheimer's disease as the aetiology of their clinical symptoms. In the new National Institute on Aging and Alzheimer Association criteria (Albert 2011), the presence of abnormally high t-tau/p-tau or low ABeta is thought to indicate MCI due to Alzheimer's disease. This is not supported by this or previous reports. What is more consistent with our findings is that, in the presence of normal levels of CSF t-tau/p-tau and ABeta, we can say MCI is not due to Alzheimer's disease. This is an important distinction and one that has important implications when conveying risk of progression to people with MCI, as well as when giving pretest counselling to people before the test takes place. The positive and negative likelihood ratios we have generated as illustrations of our findings do demonstrate that positive and negative tests do have a small change from pre- to post-test probability. However, they are relatively small and way below what would be expected from standard thresholds for a good test. The language used in new criteria do not reflect this level of uncertainty or small incremental benefit, and therefore confer much greater diagnostic accuracy on these tests than is currently merited. Our review suggests that where these tests are used to assist clinical diagnosis, their limitations and low incremental benefit should be considered. In the absence currently of any disease modifying interventions, the risk of overdiagnosis to a patient may do greater harm than underdiagnosis. However, this is a rapidly moving field and if disease-modifying or secondary prevention interventions become available, then this opinion will shift, more so if the interventions are effective, low cost and well tolerated.

Implications for research

These tests though still have value in clinical trials with drugs proposed to affect the Alzheimer's disease process where normal levels should be used to exclude subjects from the trial with the knowledge that many individuals with 'positive' tests who are entered in to the study will still not progress to Alzheimer's disease dementia. Moreover, there may well be an interaction between test results, diagnosis, and stage of illness. For instance, abnormalities in these tests may be more specific if noted in younger people with no or minimal symptoms as opposed to older, symptomatic people where they may be reflections of nonspecific neurodegeneration, ageing or physiological reactions to ABeta oligomerisation (plaque formation). Only by undertaking longitudinal studies in mid-life, preclinical populations can we answer that proposition; these types of studies are ongoing. It is also the case that several initiatives to collate data from across numerous cohort studies are commencing. These include Dementia Platform UK and the IMI-funded European Medicines Informatic Framework (EMIF) (http://www.emif.eu; http://www.dementiasplatform.uk) and European Prevention of Alzheimer's Dementia (EPAD) (Ritchie 2016) programme that will deliver analyses of source-aggregated data that will, in most cases, have been collected under standardised conditions, as well as (in the case of EPAD) developing a new longitudinal cohort which will provide at least ten-fold increase in sample size for predementia disease modelling (Ritchie 2016). New cohort data are also being regularly published, especially as extant cohorts undergo further assessments. Together this suggests that as a rapidly moving field, this review will need to be updated on a regular basis and will include data from preclinical (asymptomatic) as well as MCI populations.

ACKNOWLEDGEMENTS

Obioha Ukoumunne is supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South West Peninsula. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health.

REFERENCES

References to studies included in this review

Amlien 2013 *{published and unpublished data}*

* Amlien IK, Fjeli AM, Walhovd KB, Selnes P, Stenset V, Grambaite R, at al. Mild cognitive impairment: cerebrospinal fluid tau biomarker pathologic levels and longitudinal changes in white matter integrity. *Radiology* 2013;**266**:295–303.

Buchhave 2012 *{published data only}*

* Buchhave P, Minthon L, Zetterberg H, Wallin AK, Blennow K, Hansson O. Cerebrospinal fluid levels of beta-amyloid 1-42, but not of tau, are fully changed already 5 to 10 years before the onset of Alzheimer dementia. *Archives of General Psychiatry* 2012;**69**(1):98–106.

Mattsson N, Zetterberg H, Hansson O, Andreasen N, Parnetti L, Jonsson M, et al. CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. *JAMA* 2009;**302**(4):385–93.

Eckerstrom 2010 *{published data only}*

* Eckerström C, Andreasson U, Olsson E, Rolstad S, Blennow K, Zetterberg H, et al. Combination of hippocampal volume and cerebrospinal fluid biomarkers Improves predictive value in mild cognitive impairment. *Dementia and Geriatric Cognitive Disorders* 2010;**29**: 294–300.

Mattsson N, Zetterberg H, Hansson O, Andreasen N, Parnetti L, Jonsson M, et al. CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. *JAMA* 2009;**302**(4):385–93.

Wallin A, Göthlin M, Gustavsson M, Zetterberg H. Eckerström C, Blennow K, et al. Progression from mild to pronounced MCI Is not associated with cerebrospinal fluid biomarker deviations. *Dementia and Geriatric Cognitive Disorders* 2011;**32**:193–7.

Fellgiebel 2007 *{published data only}*

* Fellgiebel A, Scheurich A, Bartenstein P, Müller MJ. FDG-PET and CSF phospho-tau for prediction of cognitive decline in mild cognitive impairment. *Psychiatry Research: Neuroimaging* 2007;**155**:167–71.

Galluzzi 2010 *{published and unpublished data}*

* Galluzzi S, Geroldi C, Ghidoni R, Paghera B, Amicucci G, Bonetti M, et al. The new Alzheimer's criteria in a naturalistic series of patients with mild cognitive impairment. *Journal of Neurology* 2010;**257**(12):2004–14.

Hampel 2004 *{published data only}*

* Hampel H, Teipel SJ, Fuchsberger T, Andreasen N, Wiltfang J, Otto M, at al. Value of CSF ß-amyloid$_{1-42}$ and tau as predictors of Alzheimer's disease in patients with mild cognitive impairment. *Molecular Psychiatry* 2004;**9**:705–10.

Hansson 2006 *{published and unpublished data}*

* Hansson O, Zetterberg H, Buchhave P, Londos E, Blennow K, Minthon L. Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: a follow-up study. *Lancet Neurology* 2006;**5**:228–34.

Mattsson N, Zetterberg H, Hansson O, Andreasen N, Parnetti L, Jonsson M, et al. CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. *JAMA* 2009;**302**(4):385–93.

Herukka 2007 *{published data only}*

Herukka SK, Hallikainen M, Soininen H, Pirttila T. CSF Ab42 and tau or phosphorylated tau and prediction of progressive mild cognitive impairment. *Neurology* 2005;**64**: 1294–7.

Herukka SK, Hallikainen M, Tervo S, Helisalmi S, Tapiola T, Soininen T, et al. Cerebrospinal fluid Aβ42, tau and phosphorylated tau predict progression in patients with cognitive impairment. *Research and Practice in Alzheimer's disease* 2006;**11**:245–50.

* Herukka SK, Helisalmi S, Hallikainen M, Tervo S, Soininen H, Pirttila T. CSF Aß42, tau and phosphorylated tau, APOE ϵ4 allele and MCI type in progressive MCI. *Neurobiology of Aging* 2007;**28**:507–14.

Herukka SK, Pennanen C, Soininen H, Pirttila T. CSF Aβ42, tau and phosphorylated tau correlate with medial temporal lobe atrophy. *Journal of Alzheimer's Disease* 2008; **14**:51–7.

Mattsson N, Zetterberg H, Hansson O, Andreasen N, Parnetti L, Jonsson M, et al. CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. *JAMA* 2009;**302**(4):385–93.

Seppala TT, Koivisto AM, Hartikainen P, Helisalmi S, Soininen H, Herukka SK. Longitudinal changes of CSF biomarkers in Alzheimer's disease. *Journal of Alzheimer's Disease* 2011;**25**:583–94.

Kester 2011 *{published data only}*

* Kester MI, Verwey NA, Van Elk EJ, Blankenstein MA, Scheltens P, Van der Flier WM. Progression from MCI to AD: predictive value of CSF Aβ42 is modified by APOE genotype. *Neurobiology of Aging* 2011;**32**(8):1372–8.

Mattsson N, Zetterberg H, Hansson O, Andreasen N, Parnetti L, Jonsson M, et al. CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. *JAMA* 2009;**302**(4):385–93.

Koivunen 2008 *{published data only}*

* Koivunen J, Pirttilä T, Kemppainen N, Aalto S, Herukka S-K, Jauhianen AM, et al. PET amyloid ligand [11C]PIB uptake and cerebrospinal fluid β-amyloid in mild cognitive Impairment. *Dementia and Geriatric Cognitive Disorders* 2008;**26**:378–83.

Monge-Argiles 2011 *{published data only}*

* Monge-Argiles JA, Munoz-Ruiz C, Pampliega-Perez A, Gomez-Lopez MJ, Sanchez-Paya J, Rodriguez-Borja E, et al. Biomarkers of Alzheimer's disease in the cerebrospinal fluid of Spanish patients with mild cognitive impairment. *Neurochemical Research* 2011;**36**:986–93.

Monge-Argiles JA, Sanchez-Payab J, Munoz-Ruiz C,

Pampliega-Perez A, Gomez-Lopez MJ, Borja ER, et al. Patients with mild cognitive impairment and a reduced CSF Aβ1-42 protein progress rapidly to Alzheimer's disease. *Neurologia* 2012;**27**(1):28–33.

Palmqvist 2012 ***{published data only}***

Hertzel J, Minthono L, Zetterberg H, Vanmechelend E, Blcnnowc K, Hanssonllb O. Evaluation of CSF biomarkers as predictors of Alzheimer's disease: a clinical follow-up study of 4.7 years. *Journal of Alzheimer's Disease* 2010;**21**: 1119–28.

Mattsson N, Zetterberg H, Hansson O, Andreasen N, Parnetti L, Jonsson M, et al. CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. *JAMA* 2009;**302**(4):385–93.

* Palmqvist S, Hertze J, Minthon L, Wattmo C, Zetterberg Z, Blennow K, at al. Comparison of brief cognitive tests and CSF biomarkers in predicting Alzheimer's disease in mild cognitive impairment: six-year follow-up study. *PLoS ONE* 2012;**7**(6):e38639. [DOI: 10.1371/journal.pone.0038639

Parnetti 2012 ***{published data only}***

* Parnetti L, Chiasserini D, Eusebi P, Giannandrea D, Bellomo G, De Carlo C, et al. Performance of Aβ1-40, Aβ1-42, total tau,and phosphorylated tau as predictors of dementia in a cohort of patients with mild cognitive impairment. *Journal of Alzheimer's Disease* 2012;**29**:229–38.

Visser 2009 ***{published and unpublished data}***

Mattsson N, Zetterberg H, Hansson O, Andreasen N, Parnetti L, Jonsson M, et al. CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. *JAMA* 2009;**302**(4):385–93.

* Visser PJ, Verhey F, Knol DL, Scheltens P, Wahlund L-O, Freund-Levi Y, et al. Prevalence and prognostic value of CSF markers of Alzheimer's disease pathology in patients with subjective cognitive impairment or mild cognitive impairment in the DESCRIPA study: a prospective cohort study. *Lancet Neurology* 2009;**8**:619–27.

Vos 2013 ***{published and unpublished data}***

* Vos SJB, Van Rossum IA, Verhey F, Knol DL, Soininen H, Wahlund L-O, et al. Prediction of Alzheimer disease in subjects with amnestic and nonamnestic MCI. *Neurology* 2013;**80**:1–9.

References to studies excluded from this review

Desikan 2011 ***{published data only}***

* Desikan RS, McEvoy LK, Thompson WK, Holland D, Roddey JC, Blennow K, et al. Amyloid-β associated volume loss occurs only in the presence of phospho-tau. *Annals of Neurology* 2011;**70**:657–61.

Forlenza 2010 ***{published data only}***

* Forlenza OV, Diniz BS, Talib LL, Radanovic M, Yassuda MS, Ojopi EB, et al. Clinical and biological predictors of Alzheimer's disease inpatients with amnestic mild cognitive impairment. *Revista Brasileira de Psiquiatria* 2010;**32**(3): 216–22.

Holland 2012 ***{published data only}***

* Holland D, Desikan RS, Dale AM, McEvoy LK, Alzheimer's Disease Neuroimaging Initiative. Rates of decline in Alzheimer Disease decrease with age. *PLOS One* 2012;**7**(8):e42325. [DOI: 10.1371/journal.pone.0042325

Ivanoiu 2005 ***{published data only}***

* Ivanoiu A, Sindic CJ. Cerebrospinal fluid tau protein and amyloid β42 in mild cognitive impairment: prediction of progression to Alzheimer's disease and correlation with the neuropsychological examination. *Neurocase* 2005;**11**(1): 32–9.

Jack 2011 ***{published data only}***

* Jack CR, Vemuri P, Wiste BA, Weigand SD, Aisen PS, Trojanowski J, et al. Evidence for ordering of Alzheimer disease biomarkers. *Archives of Neurology* 2011;**68**(12): 1526–35.

Jagust 2009 ***{published data only}***

* Jagust WJ, Landau SM, Shaw M, Trojanowski JQ, Koeppe RA, Reiman EM, et al. Relationships between biomarkers in aging and dementia. *American Academy of Neurology* 2009;**73**:1193–9.

Lanari 2009 ***{published data only}***

* Lanari A, Parnetti L. Cerebrospinal fluid biomarkers and prediction of conversion in patients with mild cognitive Impairment: 4-year follow-up in a routine clinical setting. *Scientific World* 2009;**9**:961–6.

Landau 2010 ***{published data only}***

Davatzikosa C, Bhatta P, Shawb LM, Batmanghelich KN, Trojanowski JQ. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiology of Aging* 2011;**32**:2322.e19–2322.e27.

Ewers M, Schmitz S, Hansson O, Walsh C, Fitzpatrick A, Bennett D, et al. Body mass index is associated with biological CSF markers of core brain pathology of Alzheimer's disease. *Neurobiology of Aging* 2012;**33**: 1599–608.

Ewers M, Walsh C, Trojanowski JQ, Shaw LM, Petersen RC, Jack Jr. CR, et al. Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. *Neurobiology of Aging* 2012;**33**:1203–14.

Gomar JJ, Bobes-Bascaran MT, Conejero-Goldberg C, Davies P, Goldberg TE. Utility of combinations of biomarkers, cognitive markers, and risk factors to predict conversion from mild cognitive impairment to Alzheimer disease in patients in the Alzheimer's Disease Neuroimaging Initiative. *Archives of General Psychiatry* 2011;**68**(9):961–9.

* Landau SM, Harvey D, Madison CM, Reiman EM, Foster NL, Aisen PS, et al. Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology* 2010; **75**:230–8.

Maruyama 2004 ***{published data only}***

* Maruyama M, Matsui T, Tanji H, Nemoto M, Tomita N, Ootsuki M, et al. Cerebrospinal fluid tau protein and periventricular white matter lesions in patients with mild cognitive impairment. *Archives of Neurology* 2004;**61**: 716–20.

Maruyama 2004b *{published data only}*

* Maruyama M, Matsui T, Tanji H, Ootsuki M, Nemoto M, Tomita N, et al. Diagnosing the mild cognitive impairment stage of Alzheimer's disease. *Psychiatria et Neurologia Japonica* 2004;**106**(3):269–80.

Mattsson 2012 *{published data only}*

* Mattsson N, Porteliusa E, Rolstad S, Gustavsson M, Andreasson U, Stridsberg M, et al. Longitudinal cerebrospinal fluid biomarkers over four years in mild cognitive impairment. *Journal of Alzheimer's Disease* 2012; **30**:767–78.

Nordlund 2010 *{published data only}*

* Nordlund A, Rolstad S, Klang O, Edman A, Hansen S, Wallin A. Two-year outcome of MCI subtypes and aetiologies in the Goteborg MCI study. *Journal of Neurology, Neurosurgery, and Psychiatry* 2010;**81**:541–6.

Okamura 2002 *{published data only}*

* Okamura N, Arai H, Maruyama M, Higuchi M, Matsui T, Tanji H, et al. Combined analysis of CSF tau levels and [123I]Iodoamphetamine SPECT in mild cognitive impairment: implications for a novel predictor of Alzheimer's Disease. *American Journal of Psychiatry* 2002; **159**:474-6.

Okonkwo 2011 *{published data only}*

* Okonkwo OC, Mielke MM, Griffith HR, Moghekar AR, O'Brien RJ, Shaw LM, et al. Cerebrospinal fluid profiles and prospective course and outcome in patients with amnestic mild cognitive impairment. *Archives of Neurology* 2011;**68**(1):113–9.

Pereira 2010 *{published data only}*

* Pereira FS, Yassuda MS, Olivira AM, Diniz BS, Radanovic M, Talib LL, et al. Profiles of functional deficits in mild cognitive impairment and dementia: benefits from objective measurement. *International Neuropsychological Society* 2010; **16**:297–305.

Perneczky 2011 *{published data only}*

* Perneczky R, Tsolakidou A, Arnold A, Diehl-Schmid J, Grimmer T, Forstl H, et al. CSF soluble amyloid precursor proteins in the diagnosis of incipient Alzheimer disease. *Neurology* 2011;**77**:35–8.

Riemenschneider 2002 *{published data only}*

* Riemenschneider M, Lautenschlager N, Wagenpfeil S, Diehl J, Drzezga A, Kurz A. Cerebrospinal fluid tau and beta-amyloid 42 proteins identify Alzheimer disease in subjects with mild cognitive impairment. *Archives of Neurology* 2002;**59**(11):1729–34.

Samtani 2012 *{published data only}*

* Samtani MN, Raghavan N, Shi Y, Novak G, Farnum M, Lobanov V, et al. Disease progression model in subjects with mild cognitive impairment from the Alzheimer's disease neuroimaging initiative: CSF biomarkers predict population subtypes. *British Journal of Clinical Pharmacology* 2012;**75** (1):146–61.

Schneider 2010 *{published data only}*

* Schneider LS, Kennedy RE, Cutter GR, Alzheimer's Disease Neuroimaging Initiative. Requiring an amyloid-b1-42 biomarker for prodromal Alzheimer's disease or mild cognitive impairment does not lead to more efficient clinical trials. *Alzheimers & Dementia* 2010;**6**:e367–77.

Shaw 2009 *{published data only}*

* Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Annals of Neurology* 2009;**65**:403–13.

Sluimer 2010 *{published data only}*

* Sluimer JD, Bouwman FH, Vrenken H, Blankenstein MA, Barkhof F, Van der Flier WM, et al. Whole-brain atrophy rate and CSF biomarker levels in MCI and AD: a longitudinal study. *Neurobiology of Aging* 2010;**31**:758–64.

Snider 2009 *{published data only}*

* Snider BJ, Fagan AM, Roe C, Shah AR, Grant EA, Xiong C. Cerebrospinal fluid biomarkers and rate of cognitive decline in very mild dementia of the Alzheimer type. *Archives of Neurology* 2009;**66**(5):638–45.

Van Harten 2012 *{published data only}*

* Van Harten AC, Visser PJ, Pijnenburg YAL, Teunissen CE, Blankenstein MA, Scheltens P, et al. Cerebrospinal fluid Aß42 is the best predictor of clinical progression in patients with subjective complaints. *Alzheimers & Dementia* 2013;**9**:481–7.

Verwey 2008 *{published data only}*

* Verwey NA, Bouwman FH, Van der Flier WM, Veerhuis R, Scheltens P, Blankenstein MA. Variability in longitudinal cerebrospinal fluid tau and phosphorylated tau measurements. *Clinical Chemistry and Laboratory Medicine* 2008;**46**(9):1300–4.

Walhovd 2010 *{published data only}*

* Walhovd KB, Fjell AM, Brewer J, McEvoy LK, Fennema-Notestine C, Hagler DJ Jr, et al. Combining MR imaging, positron-emissiontomography, and CSF biomarkers in the diagnosis and prognosis of Alzheimer disease. *American Journal of Neuroradiology* 2010;**31**(2):347–54.

Wang 2012 *{published data only}*

* Wang L, Fagan AM, Shah AR, Beg MF, Csernansky JG, Morris JC, et al. Cerebrospinal fluid proteins predict longitudinal hippocampal degeneration in early-stage dementia of the Alzheimer type. *Alzheimer Disease and Associated Disorders* 2012;**26**:314–21.

Westman 2012 *{published data only}*

* Westman E, Muehlboeck J-S, Simmons A. Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. *Neurolmage* 2012;**62**:229–38.

Yang 2012 *{published data only}*

* Yang X, Tan MZ, Qiu A. CSF and brain structural imaging markers of the Alzheimer's pathological cascade. *PLOS One* 2012;**7**(12):e47406.

References to studies awaiting assessment

Balasa 2014 *{published data only}*

Balasa M, Sanchez-Valle R, Antonell A, Bosch B, Olives J, Rami L, et al. Usefulness of biomarkers in the diagnosis and prognosis of early-onset cognitive impairment. *Journal of Alzheimer's Disease* 2014;**40**(4):919–27.

Eckerstrom 2015 *{published data only}*

Eckerström C, Olsson E, Klasson N, Berge J, Nordlund A, Bjerke M, et al. Multimodal prediction of dementia with up to 10 years follow up: the Gothenburg MCI study. *Journal of Alzheimer's disease* 2015;**44**(1):205–14.

Ewers 2012 *{published data only}*

Ewers M, Walsh C, Trojanowski JQ, Shaw LM, Petersen RC, Jack CR Jr, et al. North American Alzheimer's Disease Neuroimaging Initiative (ADNI). Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. *Neurobiology of Aging* 2012;**33**(7): 1203–14.

Leuzy 2015 *{published data only}*

Leuzy A, Carter SF, Chiotis K, Almkvist O, Wall A, Nordberg A. Concordance and diagnostic accuracy of [11C]PIB PET and cerebrospinal fluid biomarkers in a sample of patients with mild cognitive impairment and Alzheimer's disease. *Journal of Alzheimer's Disease* 2015;**45** (4):1077–88.

Additional references

Albert 2011

Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendation from the National Institute on Ageing and Alzheimer's Association workgroup. *Alzheimers & Dementia* 2011;**7**(3): 270–9.

Arevalo-Rodriguez 2015

Arevalo-Rodriguez I, Smailagic N, Roqué i Figuls M, Ciapponi A, Sanchez-Perez E, Giannakou A, et al. Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database of Systematic Reviews* 2015, Issue 3. [DOI: 10.1002/ 14651858.CD010783.pub2

Beckett 2010

Beckett LA, Harvey DJ, Gamst A, Donohueb M, Kornak J, Zhang H, et al. The Alzheimer's Disease Neuroimaging Initiative: annual change in biomarkers and clinical outcomes. *Alzheimers & Dementia* 2010;**6**:257–64.

Beynon 2013

Beynon R, Leeflang MM, McDonald S, Eisinga A, Mitchell RL, Whiting P, et al. Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE. *Cochrane Database of Systematic Reviews* 2013, Issue 9. [DOI: 10.1002/14651858.MR000022.pub3

Billingsley 1997

Billingsley ML, Kincaid RL. Regulated phosphorylation and dephosphorylation of tau protein: effects on microtubule interaction, intracellular trafficking and neurodegeneration. *Biochemical Journal* 1997;**323**(3):577–91.

Blennow 2003

Blennow K, Hampel H. CSF markers for incipient Alzheimer's disease. Lancet Neurology 2003; Vol. 2, issue 10:605–13.

Bossuyt 2008

Bossuyt PM, Leeflang MM. Chapter 6: Developing criteria for including studies. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy 2008:http:/ /methods.cochrane.org/sdt/.

Bourne 2007

Bourne J. Improving services and support for people with dementia. National Audit Office 2007.

Boxer 2005

Boxer A, Miller B. Clinical features of frontotemporal dementia. *Alzheimer Disease & Associated Disorders* 2005; **19**:S3–S6.

Bruscoli 2004

Bruscoli M, Lovestone S. Is MCI really just early dementia? A systematic review of conversion studies. *International Psychogeriatrics* 2004;**16**:129–40.

Davison 2003

Davison AC. *Statistical Models*. Cambridge University Press, 2003.

DSMIII 1987

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Washington, DC: American Psychiatric Association 1987. 3rd revised edition.

DSMIV 1994

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Washington, DC: American Psychiatric Association 1994. 4th edition.

Dubois 2007

Dubois B, Feldman HH, Jacova C, DeKosky ST, Barberger-Gateau P, Cummings J, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurology* 2007;**6**(8):734–6.

Dubois 2010

Dubois B, Feldman HH, Jacova C, Cummings JL, Dekosky ST, Barberger-Gateau P, et al. Revising the definition of Alzheimer's disease: a new lexicon. *Lancet Neurology* 2010;**9** (11):1118–27.

Dubois 2014

Dubois B, Feldman HH, Jacova C, Hampel H, Molinuevo L, Blenonow K, et al. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *Lancet Neurology* 2014;**13**:614–29.

Ferri 2005

Ferri CP, Prince M, Brayne C, Brodaty H, Fratiglioni L, Ganguli M, et al. Global prevalence of dementia: a Delphi consensus study. *Lancet* 2005;**366**(9503):2112–7.

Geslani 2005

Geslani D, Tierney M, Herrmann N, Szalai J. Mild cognitive impairment: an operational definition and its

conversion rate to Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders* 2005;**19**:383–9.

Knapp 2007
Knapp M, Prince M. Dementia UK: A report to the Alzheimer's Society on the prevalence and economic cost of dementia in the UK produced by King's College London and London School of Economics. London (UK): Alzheimer's Society; 2007.

Knopman 2001
Knopman D, DeKosky S, Cummings J, Chui H, Corey-Bloom J, Relkin N, et al. Practice parameter: diagnosis of dementia (an evidence-based review): report of the quality standards subcommittee of the American Academy of Neurology. *Neurology* 2001;**56**(9):1143–53.

Knottnerus 2002
Knottnerus JA, van Weel C, Muris JW. Evaluation of diagnostic procedures. *BMJ* 2002;**324**(7335):477–80.

LMG 1994
Lund and Manchester Groups. Clinical and neuropathological criteria for frontotemporal dementia. *Journal of Neurology, Neurosurgery, and Psychiatry* 1994;**57**(4):416–8.

Mandelkow 1998
Mandelkow EM, Mandelkow E. Tau in Alzheimer's disease. *Trends in Cell Biology* 1998;**8**:4257.

Mason 2010
Mason SE, McShane R, Ritchie CW. Diagnostic tests for Alzheimer's disease: rationale, methodology, and challenges. *International Journal of Alzheimer's Disease* 2010;**2010**: Article ID 972685.

Matthews 2008
Matthews FE, Stephan BC, McKeith IG, Bond J, Brayne C. Two-year progression from mild cognitive impairment to dementia: to what extent do different definitions agree? . *Journal of the American Geriatrics Society* 2008;**56**(8): 1424–33.

Matthews 2009
Matthews FE, Brayne C, Lowe J, McKeith I, Wharton SB, Ince P. Epidemiological pathology of dementia: attributable-risks at death in the Medical Research Council Cognitive Function and Ageing study. *PLOS Medicine* 2009;**6**:e1000180.

Mattsson 2009
Mattsson N, Zetterberg H, Hansson O, Andreasen N, Parnetti L, Jonsson M, et al. CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. *JAMA* 2009;**302**(4):385–93.

McKeith 1996
McKeith IG, Galasko D, Kosaka K, Perry EK, Dickson DW, Hansen LA, et al. Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): report of the consortium on DLB international workshop. Neurology 1996; Vol. 47, issue 5:1113–24.

McKeith 2005
McKeith IG, Dickson DW, Lowe J, Emre M, O'Brien JT, Feldman H, et al. Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium. *Neurology* 2005;**65**:1863–72.

McKhann 1984
McKhann GM, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. Neurology 1984; Vol. 34, issue 7:939–44.

Molinuevo 2014
Molinuevo JL, Blennow K, Dubois B, Engelborghs S, Lewczuk P, Perret-Liaudet A, et al. The clinical use of cerebrospinal fluid biomarker testing for Alzheimer's disease diagnosis: a consensus paper from the Alzheimer's Biomarkers Standardisation Initiative. *Alzheimers & Dementia* 2014;**10**:808–17.

Morris 1993
Morris JC. The clinical dementia rating (CDR): current version and scoring rules. *Neurology* 1993;**43**:2412–4.

MRC CFAS 2001
Neuropathology Group of the Medical Research Council Cognitive Function and Ageing Study (MRC CFAS). Pathological correlates of late-onset dementia in a multicentre, community-based population in England and Wales.. *Lancet* 2001;**357**(9251):169–75.

Neary 1998
Neary D, Snowden JS, Gustafson L, Passant U, Stuss D, Black S, et al. Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology* 1998; **51**:1546–54.

Noel-Storr 2013
Noel-Storr AH, Flicker L, Ritchie CW, Nguyen GH, Gupta T, Wood P, et al. Systematic review of the body of evidence for use of biomarkers in the diagnosis of dementia. *Alzheimers & Dementia* 2013;**9**(3):e96–e105.

Noel-Storr 2014
Noel-Storr AH, McCleery JM, Richard E, Ritchie C, Flicker L, Cullum S, et al. Reporting standards for studies of diagnostic test accuracy in dementia. *Neurology* 2014;**83**: 1–10.

Okello 2009
Okello A, Koivunen J, Edison P, Archer HA, Turkheimer F, Nagren K, et al. Conversion of amyloid positive and negative MCI to AD over 3 years: an 11C-PIB PET study. *Neurology* 2009;**73**(10):754–60.

Petersen 1999
Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Archives of Neurology 1999; Vol. 56:303–8.

Petersen 2004
Petersen RC. Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine* 2004;**256**(3):183–94.

Petersen 2009
Petersen RC, Roberts RO, Knopman DS, Boeve BF, Geda YE, Ivnik RJ, et al. Mild cognitive impairment: ten years later. *Archives of Neurology* 2009;**66**(12):1447–55.

Petersen 2010
Petersen R, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology* 2010;**74**(3):201–9.

Prince 2013
Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: a systematic review and metaanalysis. 2013; Vol. 9, issue 1:63–75.e2.

Quadas-2
University of Bristol. Quadas-2. www.bris.ac.uk/quadas/quadas-2 (accessed prior to 18 February 2017).

Quinn 2014
Quinn TJ, Fearon P, Noel-Storr AH, Young C, McShane R, Stott DJ. Informant Questionnaire on COgnitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within community dwelling populations. *Cochrane Database of Systematic Reviews* 2014, Issue 4. [DOI: 10.1002/14651858.CD010079.pub2

Reisberg 1982
Reisberg B, Ferris, SH, De Leon MJ, Crook T. The global deterioration scale for assessment of primary degenerative dementia. *American Journal of Psychiatry* 1982;**139**(9):1136-9.

Ritchie 2011
Ritchie C, Masters CL, Mason SE, Li QX, McShane R. Plasma and CSF Abeta for the longitudinal prediction of Alzheimer's disease dementia and other dementias in people with cognitive decline but no dementia. *Cochrane Database of Systematic Reviews* 2011, Issue 1. [DOI: 10.1002/14651858.CD008782.pub3

Ritchie 2014
Ritchie C, Smailagic N, Noel-Storr AH, Takwoingi Y, Flicker L, Mason SE, et al. Plasma and cerebrospinal fluid amyloid beta for the diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database of Systematic Reviews* 2014, Issue 6. [DOI: 10.1002/14651858.CD008782.pub4

Ritchie 2016
Ritchie CW, Molinuevo JL, Truyen L, Saltin A, Van der Geyten S, Lovestone S, European Prevention of Alzheimer's Dementia (EPAD) Consortium. Development of interventions for the secondary prevention of Alzheimer's dementia: the European Prevention of Alzheimer's Dementia (EPAD) project. *Lancet Psychiatry* 2016;**3**(2):179–86.. [DOI: 10.1016/S2215-0366(15)00454-X

Roman 1993
Roman GC, Tatemichi TK, Erkinjuntti T, Cummings JL, Masdeu JC, Garcia JH, et al. Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN international workshop. Neurology 1993; Vol. 43, issue 2:250–60.

SAS Institute 2011 [Computer program]
SAS Institute. SAS Version 9.2. Cary, NC, USA, 2011.

Savva 2009
Savva GM, Wharton SB, Ince PG, Forster G, Matthews FE, Brayne C, et al. Age, neuropathology, and dementia. *New England Journal of Medicine* 2009;**360**(22):230–9.

Shoji 1992
Shoji M, Golde TE, Ghiso J, Cheung TT, Estus S, Shaffer LM, et al. Production of the Alzheimer amyloid beta protein by normal proteolytic processing. *Science* 1992;**258**(5079):126–9.

Shoji 2001
Shoji M, Kanai M. Cerebrospinal fluid Ab40 and Ab42: natural course and clinical usefulness. *Journal of Alzheimer's Disease* 2001;**3**(3):313–21.

Shoji 2002
Shoji M, Matsubara E, Murakami T, Manabe Y, Abe K, Kanai M, et al. Cerebrospinal fluid tau in dementia disorders: a large scale multicenter study by a Japanese study group. Neurobiology of Aging 2002; Vol. 23, issue 3:363–70.

Smailagic 2015
Smailagic N, Vacante M, Hyde C, Martin S, Ukoumunne O, Sachpekidis C. ^{18}F-FDG PET for the early diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database of Systematic Reviews* 2015, Issue 1. [DOI: 10.1002/14651858.CD010632.pub2

Urakami 2004
Urakami K, Saito K, Seishima M, Nakashima K. Apolipoprotein A-I and E in cerebrospinal fluid. *Nihon Rinsho. Japanese Journal of Clinical Medicine* 2004;**62 Suppl 11**:176–8.

Visser 2006
Visser PJ, Kester A, Jolles J, Verhey F. Ten-year risk of dementia in subjects with mild cognitive impairment. *Neurology* 2006;**67**:1201–07.

Whiting 2011
Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;**155**:529–36.

Wilmo 2010
Wilmo A, Prince M. World Alzheimer Report 2010: The global economic impact of dementia. www.alz.co.uk/research/files/WorldAlzheimerReport2010.pdf (accessed prior to 17 Feb 2017).

Winbald 2004
Winblad B, Palmer K, Kivipelto M, Jelic V, Fratiglioni L, Wahlund LO, et al. Mild cognitive impairment - beyond controversies, towards a consensus: report of the international working group on mild cognitive impairment. *Journal of Internal Medicine* 2004;**256**:240–6.

World Health Organization 2010

World Health Organization. International statistical classification of diseases and related health problems (ICD-10 Version: 2010). apps.who.int/classifications/icd10/browse/2010/en (accessed 9 January 2013):1–195.

Zhang 2014

Zhang S, Smailagic N, Hyde C, Noel-Storr AH, Takwoingi Y, McShane R, at al. 11C-PIB-PET for the early diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database of Systematic Reviews* 2014, Issue 7. [DOI: 10.1002/14651858.CD010386

* *Indicates the major publication for the study*

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Amlien 2013

Study characteristics	
Patient sampling	Study design: nested case-control study with a delayed verification design Prospective MCI group of 49 participants with MCI, who attended a university-based memory clinic, was recruited consecutively between 2005 and 2009. Twenty-three control subjects were also recruited. We only included data on performance of the index test to discriminate between participants with MCI who converted to dementia and those who remained stable Exclusion criteria: psychiatric disorder, anoxic brain damage, cancer, drug abuse, or cognitive symptoms related to solvent exposure
Patient characteristics and setting	39 MCI participants diagnosed by the Petersen 1991 criteria were included in the analysis. Ten MCI participants did not have a follow-up assessment GENDER: 20 men; 19 women AGE (y): MCI with abnormal CSF t-tau level: 64 (range 45 to 76); MC with normal CSF t-tau level: 58.5 (45 to 77) APOE ϵ4 carrier (%): not reported MMSE: MCI with abnormal CSF t-tau level: 27.2 (range 25 to 29); MC with normal CSF t-tau level: 27.9 (23 to 30) Education (y): MCI with abnormal CSF t-tau level: 12.2 (range 7 to 18); MC with normal CSF t-tau level: 11.8 (8 to 16) Sources of referral: not reported Sources of recruitment: a university-based memory clinic, Oslo, Norway.
Index tests	CSF t-tau Participants underwent lumbar puncture as part of the clinical evaluation. The CSF samples were examined for total tau levels with commercially available kits (Innogenetics, Ghent, Belgium) Threshold: This was prespecified. The age-dependent criteria for pathologic values were based on a large sample of healthy control subjects and were as follows: total tau of 300 ng/L or higher for age younger than 50 years, total tau of 450 ng/L or higher for age 50 to 69 years, and total tau of 500 ng/L or higher for age older than 70 years (Sjogren 2001). The 0.90 fractile was estimated to establish reference values for CSF t-tau Not reported whether the index test results were interpreted without knowledge of the results of the reference standard
Target condition and reference standard(s)	Target condition: conversion from MCI to Alzheimer's disease dementia Reference standard: The Global Deterioration Scale (Reisberg 1982) in combination with the research criteria for the diagnosis of Alzheimer's disease, International Working group (Dubois 2007) Not reported whether clinicians conducting follow-up were aware of CSF t-tau results
Flow and timing	Duration of follow-up: mean 2.6 ± 0.54 years At baseline: 49 MCI participants had CSF sample At follow-up: 39 participants: 9 MCI with abnormal baseline CF t-tau: 5 MCI converters and 4 MCI nonconverters; 30 MCI with normal baseline CSF t-tau: 4 MCI converters and 26 MCI nonconverters (information from the author) Number included in analysis (N=39)

	Conversion to ADD: TP = 5; FP = 4; FN = 4; TN = 26 sensitivity = 55%; specificity = 87% (calculated in Revman5) Loss to follow-up N = 17 participants (10 MCI and 7 controls) did not have a follow-up assessment. Between baseline and follow-up, four participants with MCI objected to re-examination, one died of unrelated causes, and five were excluded because of definite other diagnoses. Seven control subjects objected to re-examination. All 39 MCI participants with the follow-up assessment were included in the analysis (page 297)
Comparative	
Notes	The trial investigators contacted; they provided relevant data for creating 2 X 2 table (email on 13/12/13)

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
		Unclear	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		

Amlien 2013 (*Continued*)

Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Unclear**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
		Unclear	

Buchhave 2012

Study characteristics	
Patient sampling	Study design: prospective cohort study The inclusion of 137 participants was described in the paper by Hansson 2006, as below: 180 participants with MCI, who had sought medical advice for subjective memory difficulties, were consecutively recruited at a university hospital between July 1998 and June. Thirty-six healthy controls were also included, but appeared to have been used for comparison purposes only and not included in the ROC analysis. CSF was obtained at baseline from 137 MCI participants. Of the 43 participants with MCI who did not undergo successful lumbar puncture at baseline, 32 preferred not to go through the procedure and in 11 the procedure did not deliver usable CSF Participants with other causes of cognitive impairment, including brain tumour, subdural haematoma, CNS infection, and current alcohol abuse, were excluded Buchhave 2012 and Hansson 2006 studies used the same cohort.
Patient characteristics and setting	137 participants diagnosed by the Petersen 1991 criteria. Baseline demographic data reported for 134 participants GENDER: 60 men; 74 women AGE (y): MCI-MCI (stable) 61.9 ± 8.5; MCI-AD 73.9 ± 5.8; MCI-other dementia 71.1 ± 9.1 APOE ϵ4 carrier (%): MCI-MCI (stable) 19 (46); MCI-AD 53 (74); MCI-other dementia 5 (24) MMSE: MCI-MCI 27.5 ± 2.0; MCI-AD 26.9 ± 1.4; MCI-other dementia 26.8.0 ± 1.2 Education: not reported Sources of referral: not reported Sources of recruitment: secondary care, outpatients of memory disorder clinic, University hospital, Malmo, Sweden

Buchhave 2012 (*Continued*)

Index tests	CSF p-tau/ABeta ratio Cerebrospinal fluid was collected in polypropylene tubes, stored at -80 °C, and analysed after clinical follow-up of the study was completed. No further details Threshold: determined at follow-up (page 102); a mixture model was used to establish accurate cutoff value: used to identify optimal cut-offs: < 6.16
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia or other forms of dementia Reference standards: NINCDS-ADRDA and DSM-III-R for Alzheimer's disease dementia; DSM-III-R for vascular dementia; McKeith for Lewy bodies dementia and Brun for frontotemporal dementia Clinicians conducting follow-up were blinded to CSF biomarker results
Flow and timing	Duration of follow-up: median 9.2 years (range: 4.1 years to 11.8 years) At baseline: 137 MCI participants At follow-up:134 MCI: 72 MCI-AD; 21 MCI-other dementias; 41 MCI-MCI (stable); (page 99) Number included in analysis: 134 Conversion to Alzheimer's disease dementia: sensitivity 88%; specificity 90% (page 102) TP = 63; FP = 6; FN = 9; TN = 56 (calculated in RevMan5) Loss to follow-up: 3 participants died before completion of 4 years follow-up and were excluded from the analyses because their cognitive ability was uncertain
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		

If a threshold was used, was it pre-specified?	No		
		High	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Low	

Eckerstrom 2010

Study characteristics	
Patient sampling	Study design: retrospective analysis of the longitudinal data Retrospective recruitment of 42 participants with MCI from the Gothenborg study: 21 MCI converters and 21 MCI-stable participants. The group of MCI converters comprised all MCI converters who underwent a baseline MRI investigation. The MCI-stable participants were included consecutively to achieve matching group size. No further details Twenty-six controls were also recruited. We only included data on performance of the index test to discriminate between participants with MCI who converted to dementia and those who remained stable Exclusion criteria: not reported.
Patient characteristics and setting	42 participants, diagnosed with the Global Deterioration Scale (GDS) criteria (Reisberg 1998) at baseline. GDS criteria were not prespecified diagnostic criteria for MCI

	Gender: 18 male; 24 female Age: total sample: mean age 67.9 (range 51 to 78) years; MCI-MCI: 66.6 (range 56 to 78) years; MCI-progressive: 69.3 (range 51 to 78) years APOE ε 4: not reported MMSE: mean 27.8 (range 22 to 30); MCI-MCI: 28.3 (range 24 to 30); MCI-progressive: 27.2 (range 22 to 30) Education: total sample: mean age 11.4 (range 6 to 19) years; MCI-MCI: 12.5 (range 8 to 19) years; MCI-progressive: 10.4 (range 6 to 17.5) years Sources of referral: not reported Sources of recruitment: University of Gothenburg, Molndal, Sweden
Index tests	CSF t-tau CSF samples were collected by lumbar puncture (LP). Both baseline and follow-up lumbar punctures were performed in the morning to exclude influences on the results from possible diurnal fluctuations in biomarker levels. The samples were collected in polypropylene tubes that were stored at -80 °C, without being thawed and re-frozen, pending biochemical analyses. All CSF analyses for a participant were performed on the same occasion. CSF levels were determined using a sandwich enzyme-linked immunosorbent assay constructed to measure tau Threshold: 500 ng/L, not prespecified. ROC curves were used to calculate the cutoff values based on the maximum for the sum of sensitivity and specificity It was not reported whether the index test results were interpreted without knowledge of the results of the reference standard
Target condition and reference standard(s)	Target condition: conversion from MCI to ADD or FTD or subcortical VD or mixed AD/VD Reference standard: NINCDS-ADRDA criteria; Lund and Manchester criteria; Erkinjunitti criteria In addition, Global Deterioration Scale (GDS) criteria (Reisberg 1998): score = 4 were used It was not reported whether the results of the reference standard were interpreted without knowledge of the results of the index test
Flow and timing	Duration of follow-up: 2 years At baseline: 42 MCI participants At follow-up: 21 MCI converters (13 MCI-AD; 4 MCI-VD; 2 MCI-FTD; 2 MCI-AD/VD); 21 MCI- stable (MCI-MCI) (page 296) 21 MCI converters and 21 MCI non-converters were selected from the Gothenborg study for the retrospective analysis Number included in analysis: 42 Conversion to all types of dementia: sensitivity 67%; specificity 95% (Table 3, page 298) TP = 14; FP = 1; FN = 7; TN = 20 (Fig 1, page 297) Insufficient data to create 2 X 2 table for conversion from MCI to ADD Loss to follow-up: all retrospectively selected participants were included in the analysis
Comparative	
Notes	
Methodological quality	

Eckerstrom 2010 (*Continued*)

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
		High	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
		High	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Unclear**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		

Eckerstrom 2010 (*Continued*)

Were all patients included in the analysis?	Unclear		
		Unclear	

Fellgiebel 2007

Study characteristics	
Patient sampling	Study design: prospective cohort study Prospective recruitment of 16 participants with amnestic mild cognitive impairment, presenting at a memory clinic for diagnostic evaluation. Sampling procedure not described Exclusion criteria: participants with metabolic disease that could affect cognitive function; participants with other brain diseases; participants with a diagnosis of depression according to DSM-IV criteria
Patient characteristics and setting	16 participants, diagnosed with the Petersen 1999 criteria at baseline. One participant of the initial study group refused further participation and was replaced by a consecutively recruited comparable participant of the memory clinic to preserve the statistical power for prospectively planned follow-up analyses Gender: 9 male; 7 female Age: total sample: mean age 68.6 ± 7.9 years; MCI-MCI: 68.8 ± 10.0 years; MCI-progressive: 68.5 ± 5.9 years (4/8 MCI-AD: 69.5 ± 7.9 years) APOE ε4: not reported MMSE: mean 25.7 ± 2.7; MCI-MCI: 27.3 ± 1.8; MCI-progressive: 25.0 ± 2.1 (4/8 MCI-AD: 24.3 ± 1.5) Education: not reported Sources of referral: not reported Sources of recruitment: University Memory Clinic, Germany
Index tests	CSF p-tau Method of the index test administration described previously (Fellgiebel 2004). CSF was analysed with two sandwich enzyme-linked immunosorbent assays: Tau protein phosphorylated at threonine 181 (p-tau181) was determined using the Innogenetics INNOTEST Phospho-Tau(181) kit and total tau protein (t-tau) was examined with the INNOTEST-hTau-Ag kit Threshold: Besides the previous published p-tau181 cutoff (Fellgiebel 2004), the cutoff value of 50 pg/mL was chosen as an optimal cutoff by means of Receiver Operating Characteristic (ROC) analyses to separate participants with MCI from controls (measures in 75 participants, unpublished data) Not reported whether the index test results were interpreted without knowledge of the results of the reference standard
Target condition and reference standard(s)	Target condition: conversion from MCI to Alzheimer's disease dementia Reference standard: progression to Alzheimer's disease dementia was assumed if CDR reached 1 Follow-up evaluation at variable time points (not specified), comprising neurological and psychiatric examination, CDR and MMSE Progressive cognitive decline was defined as MMSE score reduction >/= 2 and a clinical judgement of cognitive deterioration

	Clinicians conducting follow-up were blinded to the CSF p-tau results
Flow and timing	Duration of follow-up: total sample: 19.6 ± 9.0 months; MCI-MCI: 19.5 ± 9.3 months; MCI-progressive: 17.6 ± 8.8 months (4/8 MCI-AD: 23.7 ± 2.0 months) At baseline: 16 MCI: 12 CSF p-tau positive; 4 CSF p-tau negative. At follow-up: 16 MCI: 12 CSF positive: 4 MCI-AD (converters), 8 MCI-MCI (non-converters) , 4 MCI-progressive (non-converters); 4 CSF p-tau negative: 4 MCI-MCI (stable non-converters) (page 170) Number included in analysis: 16 Conversion to AD: TP = 4; FP = 8; FN = 0; TN = 4 sensitivity: 100%; specificity: 33% (calculated in RevMan5) Loss to follow-up:1/16; however, that participant was replaced by an additional, consecutively recruited patient from the memory clinic
Comparative	
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Unclear	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
		High	**Low**
DOMAIN 3: Reference Standard			

Fellgiebel 2007 (*Continued*)

Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Unclear	**Unclear**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Unclear	

Galluzzi 2010

Study characteristics	
Patient sampling	Study design: retrospective analysis of the longitudinal data. Retrospective MCI group of 108 participants was recruited consecutively in a previous prospective study from the university memory clinic over 24 months Participants were excluded if they had a history or presence of neurological signs of major stroke
Patient characteristics and setting	90 participants, who had been diagnosed by the Petersen 1999 criteria at baseline and had follow-up assessment, were included in the study. The remaining 18 participants were not included in the study because they lacked follow-up assessment due to refusal (n = 16) or logistic problems (n = 2). CSF was obtained only from 64 participants. Demographic data were reported on 90 participants Gender: 37 men; 53 women Age: mean 72.05 years; MCI-MCI: 70.09 ± 7.1; MCI-AD: 72.2 ± 7.1; MCI-nAD: 25.5 ± 1.9 APOE ϵ4 carrier: 35; MCI-MCI: 19; MCI-AD: 14; MCI-nAD: 2 MMSE: MCI-MCI: 26.3 ± 1.9; MCI-AD: 26.4 ± 1.6; MCI-other dementia: 73.0 ± 7.1 Sources of recruitment: secondary care, outpatients from Translational Outpatient Memory Clinic (TOMC), Brescia, Italy
Index tests	CSF t-tau CSF was obtained by lumbar tap between L4 and L5 or L3 and L4 and processed, as detailed elsewhere (Frisoni 2009). Levels of CSF proteins were determined by commercially available enzyme-

	linked immunosorbent assay (Innogenetics, Belgium) Threshold: > 450 pg/mL for subjects with an age range between 51 and 70 years determined; > 500 pg/mL for subjects with an age range between 71 and 93 years; threshold determined at baseline and based on published criteria (page 2006) Not reported whether the index test results were interpreted without knowledge of the results of the reference standard
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia or other forms of dementia Reference standard: NINCDS-ADRDA criteria Unclear whether clinicians conducting follow-up were aware of initial CSF analysis results
Flow and timing	Duration of follow-up: 24.0 ± 9.7 months At follow-up: 39/90 participants converted to dementia (Abstract) Number included in analyses: 64 24 MCI with 'abnormal CSF t-tau level': 19 MCI converters and 5 MCI-MCI; 40 MCI with 'normal CSF t-tau level': 15 MCI converters and 25 MCI-MCI (from Dr Galluzzi's email) Conversion to all forms of dementia: TP = 19; FP = 5; FN = 15; TN = 25 sensitivity = 56%; specificity = 83% Loss to follow-up: 26 (24 participants refused the LP procedure; 2 LPs were not performed due to osteoarthrosis)
Comparative	
Notes	The trial investigators were contacted; they provided data tor the 2 X 2 table to be completed for conversion to all forms of dementia. Normative data for CSF p-tau and CSF t-tau/ABeta ratio were not available (email from Dr Galluzi on 9/12/13)

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Low	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference stan-	Unclear		

Galluzzi 2010 (*Continued*)

dard?			
If a threshold was used, was it pre-specified?	Yes		
		Unclear	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
		High	

Hampel 2004

Study characteristics	
Patient sampling	Study design: retrospective analysis of the longitudinal data. The MCI group was not a consecutive sample. Retrospective recruitment of 52 participants with MCI: 29 MCI converters and 23 MCI-stable participants. In addition, 93 participants with probable AD and 10 healthy, age-matched controls were recruited from a hospital rehabilitation department. We only included data on performance of the index test to discriminate between participants with MCI who converted to dementia and those who remained stable.The MCI group was not a consecutive sample and was selected at follow-up No exclusion criteria were specified.

Hampel 2004 (*Continued*)

Patient characteristics and setting	52 MCI participants diagnosed by the Petersen 1991 criteria at baseline Gender: 24 men; 28 women Age: mean age 72.6 years (range 54 to 87) APOE ϵ4 carrier: not reported MMSE (all MCI): 28.9 ± 1 (range 26 to 30) Sources of referral: not reported Sources of recruitment: secondary care, inpatients from Department of Rehabilitation, Pitea, Sweden
Index tests	CSF t-tau biomarker CSF samples were taken by lumbar puncture, collected in polypropylene tubes, and stored at -80 ºC. T-tau was measured in duplicate using a commercial ELISA (Innotest beta-amyloid 1-42, Innogenetics, Belgium) Threshold(s): ≥ 479 ng/L, established in the MCI-MCI vs MCI-AD at follow-up (page 707) At follow-up: 14 with 'normal CSF t-tau level' and 38 with 'abnormal CSF t-tau level' (calculated in RevMan5) Unclear whether the index test results were interpreted without knowledge of the results of the reference standard
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia Reference standard: NINCDS-ADRDA criteria; DSM-IV criteria. All MCI participants were assessed with both reference standards Unclear whether clinicians conducting follow-up a were aware of initial CSF analysis results
Flow and timing	Duration of follow-up: mean 8.4 ± 5.1 months (range 2 to 24 months); follow-up interval for converters was 9.6 ± 5.4, and for non-converters 7.0 ± 4.3 months At follow-up: 52 MCI: 29 MCI-AD; 23 MCI-MCI (page 94) Number included in analysis (N=52) Conversion to Alzheimer's disease dementia: sensitivity 90%; specificity 48% (page 707); disease positive: 29; disease negative: 23 TP = 26; FP = 12; FN = 3; TN = 11 (calculated in RevMan5) Loss to follow-up: data for all 52 MCI participants were reported
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		

Did the study avoid inappropriate exclusions?	Unclear		
		High	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
		High	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Unclear	

Hansson 2006

Study characteristics	
Patient sampling	Study design: nested case-control study with a delayed verification design Prospective cohort of 180 participants with MCI, who had sought medical advice for subjective memory difficulties, was consecutively recruited at a university hospital between July 1998 and June 2001. 39 healthy controls were also included, but appeared to have been used for comparison purposes only and not included in the ROC analysis Patients with other causes of cognitive impairment, including brain tumour, subdural haematoma, CNS infection, and current alcohol abuse, were excluded
Patient characteristics and setting	137 MCI participants, diagnosed by the Petersen 1999 and Petersen 2004 criteria, underwent successful lumbar puncture. Baseline demographic data reported for 134 participants. Of the 43 participants with MCI who did not undergo successful lumbar puncture at baseline, 32 preferred not to go through the procedure and in 11 the procedure did not deliver usable CSF GENDER: 60 men; 73 women AGE (median (range)): MCI-MCI (stable): 67 (50 to 86) years; MCI-AD 75 (59 to 85) years; MCI-other dementia 76 (54 to 82) years APOE ϵ4 carrier: MCI-MCI (stable): 28 (50%); MCI-AD 43 (75%); MCI-other dementia 6 (29%) MMSE: mean ± SD: MCI-MCI (stable) 27.3 ± 1.8; MCI-AD 26.8 ± 1.4; MCI-other dementia 27.0 ± 1.5 Education (higher): MCI-MCI (stable): 26 (46%); MCI-AD 18 (32%); MCI-other dementia 10 (48%) Sources of referral: most participants (75%) by family practitioners Sources of recruitment: secondary care, outpatients from memory disorder clinic, University hospital, Malmo, Sweden
Index tests	CSF t-tau; CSF p-tau; CSF p-tau/ABeta ratio CSF samples were obtained and stored in polypropylene tubes at -80ºC, and analysed after the clinical follow-up of the study was completed T-tau, tau phosphorylated at threonine 181 (P-tau181), and ABeta42 concentrations were measured with xMAP technology and the INNOBIA AkzBio3 kit (innogenetics), as previously described in detail (Olssson 2005). The CSF concentrations of t-tau, p-tau181, and ABeta42 were highly correlated to the concentrations obtained with conventional ELISA measurements. The best cutting values for the different combinations of the CSF biomarkers were established in the whole control and MCI patient material as those giving the highest Youden index (Youden 1950) Threshold: > 350 pg/mL for CSF t-tau; ≥ 60 pg/mL for CSF p-tau; < 6.5 pg/mL for CSF p-tau/ABeta It was not reported whether the index test results were interpreted without knowledge of the results of the reference standard
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia or other forms of dementia Reference standards: NINCDS-ADRDA and DSM-III-R for Alzheimer's disease dementia; NINDS-AIREN and DSM-III-R for vascular dementia; McKeith for Lewy bodies dementia and Brun for frontotemporal dementia Clinicians making the diagnosis during follow-up were unaware of all CSF analyses
Flow and timing	Duration of follow-up: total sample: median 5.2 years (range 4.0 to 6.8); MCI-AD: median: 4.3 years (range 1.1 to 6.7); MCI-other dementias: median 4.2 (1.5 to 6.3) **CSF t-tau** At baseline: 137 MCI participants with CSF sample

Hansson 2006 (*Continued*)

	At follow-up: 134: 38 MCI with baseline positive CSF t-tau: 29 MCI-AD; 4 MCI-other dementias; 5 MCI-MCI; 96 MCI with baseline negative CSF t-tau: 28 MCI-AD; 17 MCI-other dementias; 51 MCI-MCI Number included in analysis: 134 1) Conversion to AD: TP = 29; FP = 9; FN = 28; TN = 68; sensitivity = 51%; specificity = 88% (calculated in Revman5) 2) Conversion to all dementias: TP = 33; FP = 5; FN = 45; TN = 51; sensitivity = 42%; specificity = 91% (calculated in Revman5) **CSF p-tau** At baseline: 137 MCI participants with CSF sample At follow-up: 134: 50 CSF p-tau positive: 39 MCI-AD; 2 MCI-other dementias; 9 MCI-MCI; 84 CSF p-tau negative: 18 MCI-AD; 19 MCI-other dementias; 47 MCI-MCI Number included in analysis: 134 1) Conversion to AD: TP = 39; FP = 11; FN = 18; TN = 66; sensitivity = 68%; specificity = 86% (calculated in Revman5) 2) Conversion to all dementias: TP = 41; FP = 9; FN = 37; TN = 47; sensitivity = 52%; specificity = 84% (calculated in Revman5) **CSF p-tau/Aß ratio** At baseline: 137 MCI participants with CSF sample At follow-up: 134: 74 CSF p-tau/ABeta positive: 55 MCI-AD; 4 MCI-other dementias; 15 MCI-MCI; 60 CSF p-tau/ABeta negative: 2 MCI-AD; 17 MCI-other dementias; 41 MCI-MCI Number included in analysis: 134 1) Conversion to AD: TP = 55; FP = 19; FN = 2; TN = 58; sensitivity = 96%; specificity = 75% (calculated in Revman5) 2) Conversion to all dementias: TP = 59; FP = 15; FN = 19; TN = 41; sensitivity = 76%; specificity = 73% (calculated in Revman5) Loss to follow-up: initially identified 180 consecutive participants with MCI; 43/180 not included in the study: 32 refused lumbar puncture and 11 non-usable CSF samples; 3 participants died before 4 years of follow-up (not included in the analysis)
Comparative	
Notes	The trial investigators contacted; they provided relevant data for creating 2 X 2 table items (email on 29/11/13)

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		

Did the study avoid inappropriate exclusions?	Yes		
		Low	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
		High	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
		High	

Herukka 2007

Study characteristics	
Patient sampling	Study design: nested case-control study with delay verification design. Participants examined in a university hospital neurological department, or from an ongoing population-based study were prospectively recruited if they agreed to a lumbar puncture for research purposes and had a baseline diagnosis of MCI; 79 participants met these criteria. 60 controls (who were referred to the neurological department for different symptoms, or who were included in the population-based study and had depression with normal performance in neuropsychological tests) were also included. We only included data on performance of the index test to discriminate between participants with MCI who converted to dementia and those who remained stable No exclusion criteria were specified.
Patient characteristics and setting	79 MCI participants diagnosed by the CDR = 0.5 criteria at baseline Gender: 33 men; 46 women Age: 70.56 years; MCI-MCI: 69.46 ± 8.14; MCI-progressive: 71.76 ± 6.71 APOE ϵ4 carrier: 41; MCI-MCI: 15/45 (33.3%); MCI-progressive: 26/33 (78.8%) MMSE: MCI-MCI: 24.09 ± 2.49; MCI-progressive: 23.91 ± 2.69 Sources of recruitment: secondary care, inpatients from neurological department, Kuopio University Hospital, Finland
Index tests	CSF t-tau; CSF p-tau The CSF samples were collected by LP during the baseline visit. The samples were stored in polypropylene tubes at -70 °C until analysis. The measurement of CSF t-tau and CSF P-tau were done by using a commercial ELISA (Innogenetics, Belgium), blinded to the diagnoses Threshold: > 400 pg/mL for CSF t-tau; > 70 pg/L for CSF P-tau; thresholds determined at baseline using previously published cutoff values from the ROC analysis (Herruka 2005) Index test was conducted at baseline and interpreted blinded to the diagnoses of APOE genotype
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia or other forms of dementia Reference standard: NINCDS-ADRDA for Alzheimer's disease dementia; DSM-IV-R criteria for other dementias Diagnosis of dementia was done independently and blinded to CSF biomarker results
Flow and timing	Duration of follow-up: mean 3.52 ± 1.95 years in MCI converters; mean 4.56 ± 3.09 years in MCI-stable At follow-up:79 MCI: 33 MCI converters (27 MCI-AD; 1 MCI-SVD; 5 MCI-MD); 46 MCI-MCI (page 509) Number included in analyses = 79 Conversion from MCI to all dementias (Fig 1, page 510): 1) CSF t-tau: TP = 26, FP = 17, FN = 7, TN = 29; sensitivity = 79%; specificity = 63% (calculated in Revman5) 2) CSF P-tau:TP = 25, FP = 16, FN = 8, TN = 30; sensitivity = 76%; specificity = 65% (calculated in Revman5) Loss to follow-up: CSF marker and follow-up data appeared to have been available for all participants
Comparative	
Notes	

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
		High	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		

Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Low	

Kester 2011

Study characteristics	
Patient sampling	Study design: retrospective analysis of the longitudinal data 153 participants with available CSF results and APOE ϵ4 genotyping were diagnosed with MCI in the memory clinic in the period between January 2001 and May 2008. 107 of those 153 participants had follow-up data available and were retrospectively recruited No exclusion criteria were reported.
Patient characteristics and setting	107 MCI participants diagnosed by the Petersen 1991 criteria at baseline. Baseline demographic data reported for 100 participants, who were included in the analysis Gender: 59 men; 41 women. MCI-MCI: 38 men; 20 women; MCI-AD: 21 men; 21 women Age: 67 ± 9 years MCI-MCI; 69 ± 7 years MCI-AD APOE ϵ4 carrier: 27/58 MCI-MCI; 30/42 MCI-AD MMSE: 27 ± 2 MCI-MCI; 26 ± 3 MCI-AD Sources of referral: not reported Sources of recruitment: secondary care, outpatients from memory clinic, Amsterdam, the Netherlands
Index tests	CSF t-tau biomarker CSF was obtained by lumbar puncture between the L3/L4 and L4/L5 intervertebral space and collected in10 mL polypropylene tubes. CSFsamples were processed within 2 hours (centrifuged at 1800 × g for 10 min at 4 °C and stored at -80 °C in polypropylene tubes until analysis). CSF t-tau was measured using a commercial sandwich ELISA (Innotest) Threshold(s): > 356 pg/mL abnormal level; determined at baseline and based on published data (Schoonenboom 2005) The index test results were interpreted without knowledge of the results of the reference standard
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia or other forms of dementia Reference standard: NINCDS-ADRDA criteria for Alzheimer's disease dementia; Neary 1998 criteria; Roman 1993 criteria; McKeth 2005 criteria Not reported whether clinicians conducting follow-up were aware of CSF biomarker results
Flow and timing	Duration of follow-up: median 18 months (IQR 13 to 24); for MCI converters, the median was 17 months (IQR 13 to 24); for MCI-stable the median was 18 months (IQR 12 to 25) At follow-up: 107 MCI: 49 MCI converters (42 MCI-AD; 3 MCI-FTD, 2 MCI-VD; 1 MCI-LBD; 1 MCI-dementia due to hydrocephalus); 58 MCI-MCI (page 1373).

	Number included in analyses: 100 MCI: 42 MCI-AD and 58 MCI-MCI 36 with 'normal' CSF t-tau level and 64 with 'abnormal' CSF t-au level (Table 1, page 1374) Conversion to AD: TP = 35; FP = 29; FN = 7; TN = 29; sensitivity = 83%; specificity = 50% (calculated in Revman5) Missing data: 7 MCI participants who converted to other forms of dementia were excluded from the analysis		
Comparative			
Notes			

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	No		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
		High	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		

		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
		High	

Koivunen 2008

Study characteristics	
Patient sampling	Study design: nested case-control study with a delayed verification design Prospective MCI group of 15 participants with aMCI and 22 healthy controls were included. We only included data on performance of the index test to discriminate between participants with MCI who converted to dementia and those who remained stable. Sampling procedure and exclusion criteria not described
Patient characteristics and setting	15 participants diagnosed by the Petersen 1999 criteria. GENDER: 9 men; 6 women AGE (y): mean age 71.1 ± 7.2 APOE ϵ4 carrier (%): not reported MMSE: 25.1 (range 18 to 30) Education: not reported Sources of referral: not reported Sources of recruitment: not reported. The study was conducted in Finland.
Index tests	CSF t-tau; CSF p-tau; CSF p-tau/ABeta42 CSF sample was collected by lumbar puncture into polypropylene tubes and stored at -70 °C until analysis. The CSF levels of ABeta42, total tau and p tau (181P) were measured by a commercial ELISA (Innogenetics, Ghent, Belgium) according to the manufacturer's protocol Threshold: CSF t-tau > 400 pg/mL; CSF p-tau < 70 pg/mL; CSF p-tau/ABeta42 < 6.5 pg/mL. The cut-off values used were based on the own control material. No further information Index test was conducted before clinical follow-up. The ELISA analyses were done blinded to the diagnosis
Target condition and reference standard(s)	Target condition: conversion from MCI to Alzheimer's dementia Reference standard: NINCS-ADRDA criteria; DSM-IV criteria Not reported whether clinicians conducting follow-up were aware of CSF biomarkers' results

Flow and timing	Duration of follow-up: 2 years At baseline: 15 MCI participants (10 CSF t-tau abnormal tests; 9 CSF p-tau abnormal tests; 9 CSF p-tau/ABeta42 abnormal tests) (abstract) At follow-up: 15 MCI: 6 MCI-AD (3 CSF t-tau abnormal tests; 3 CSF p-tau abnormal tests; 4 CSF p-tau/ABeta42 abnormal tests); 9 MCI-MCI (stable) (page 381) Number included in analysis: 14 Conversion to Alzhemer's disease dementia: CSF p-tau: TP = 2; FP = 7; FN = 3; TN = 2; sensitivity = 40%; specificity = 22% (calculated in Revman5) CSF p-tau/ABeta42: TP = 4; FP = 6; FN = 1; TN = 3; sensitivity = 80%; specificity = 33% (calculated in Revman5) Loss to follow-up: CCF p-tau result was not available for one MCI-AD participant
Comparative	
Notes	The trial investigators were contacted and asked for the relevant data for CSF t-tau (email on 30/12/13). No further information was available at the time this review was prepared

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
		High	**Unclear**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			

Koivunen 2008 (*Continued*)

Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
		Unclear	

Monge-Argiles 2011

Study characteristics	
Patient sampling	Study design: nested case-control study with a delayed verification design Prospective MCI group of 37 MCI participants, attending the cognitive deterioration outpatients clinic of a general hospital, and 24 control subjects without subjective memory loss or known cognitive deterioration were recruited. No further details. We only included data on performance of the index test to discriminate between participants with MCI who converted to dementia and those who remained stable Participants with dementia or other neurological, psychiatric or medical disease which could provoke cognitive deterioration, anticoagulant therapy, failure to obtain informed consent, or a Yesavage depression scale score > 5 were excluded
Patient characteristics and setting	37 participants diagnosed by the Petersen 2006 criteria at baseline; Gender: 13 men; 24 women; MCI-MCI: 11 men, 15 women; MCI-AD: 2 men, 9 women Age: mean 73.43 ± 6.63 years APOEϵ4 carrier: not reported MMSE: mean 25 ± 2.4; MCI-AD: mean 23 ± 1.2 Sources of recruitment: secondary care, outpatients from General Hospital, Spain

Monge-Argiles 2011 (*Continued*)

Index tests	CSFt-tau; CSF p-tau; CSF t-tau/ABeta ratio; CSF p-tau/ABeta ratio The LP was performed by a hospital neurologist with a 20 X 3.5 gauge needle. CSF was collected in standard tubes and centrifuged if little sanguinolent, before being frozen. CSF samples with obvious blood were discarded. CSF biomarkers were analysed using xMAP Luminex technology and INNO-BIA Alzbio3 reagents (Innogenetics, Belgium) Threshold(s): 77.5 pg/mL for CSF t-tau; 54.5 for CSF P-tau; 0.18 for CSF t-tau/ABeta ratio; 0.17 for CSF P-tau/ABeta ratio (Table 6, page 990); thresholds determined at follow-up: ROC curve analysis was performed to determine the best cutoff values for measurement of variables. The best cutoff value was defined taking into account the highest sensitivity Index test was conducted before clinical follow-up and all samples were blindly analysed with respect to the clinical data
Target condition and reference standard(s)	Target condition: conversion to Alzheimer's disease dementia Reference standard: NINCDS-ADRDA criteria Unclear whether clinicians conducting follow-up were aware of CSF biomarker results
Flow and timing	Duration of follow-up: 6 months At baseline: 37 MCI At follow-up: 11 MCI-AD; 26 MCI-MCI (Table 1, p 989) Number included in analyses=37 Conversion to Alzheimer's disease dementia: 1) CSF t-au: sensitivity 72.7%; specificity 70% (Table 6, page 990) TP = 8, FP = 8, FN = 3, TN = 18 (calculated in RevMan5) 2) CSF p-tau: sensitivity 82%; specificity 58% (Table 6, page 990) TP = 9, FP = 11, FN = 2, TN = 15 (calculated in RevMan5) 3) CSF t-tau/ABeta ratio: sensitivity 91%; specificity 50% (Table 6, page 990) TP = 10, FP = 13, FN = 1, TN = 13 (calculated in RevMan5) 4) CSF p-tau/ABeta ratio: sensitivity 82%; specificity 66% (Table 6, page 990) TP = 9, FP = 9, FN = 2, TN = 17 (calculated in RevMan5) Loss to follow-up: CSF marker and follow-up data appeared to have been available for all participants
Comparative	
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		

Monge-Argiles 2011 (*Continued*)

		Unclear	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
		High	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Unclear	

Palmqvist 2012

Study characteristics	
Patient sampling	Study design: prospective cohort study 133 participants were "randomly recruited" among those fulfilling the MCI criteria who were referred to the memory clinic between 2000 and 2006. There were several people during this period

	who were not included due to administrative causes (information from the author) Exclusion criteria were not described.
Patient characteristics and setting	133 MCI participants, diagnosed with the Petersen 2004 criteria, were recruited from the Memory Clinic of University Hospital in Malmo, Sweden. At the initial visit, all participants were assessed by physicians experienced in dementia disorders, and underwent thorough physical, psychiatric and neurological examinations, as well as an interview that focused on their cognitive symptoms and ADL function Gender: MCI-MCI: 34 women, 28 men; MCI-AD: 36 women, 16 men; MCI-other dementias: 8 women, 11 men Age (y): MCI-MCI: mean 69.8 (range 55 to 85); MCI-AD: 75.3 (range 55 to 87); MCI-other dementias: 71.2 (59 to 83) APOEϵ4 carrier (%): MCI-MCI: 28 (45); MCI-AD: 39 (76); MCI-other dementias: 12 (63) MMSE: MCI-MCI: mean 28.1 ± 1.2; MCI-AD: mean 26.1 ± 1.5; MCI-other dementias: mean 27.1 ± 2.0 Education: not reported Sources of referral: most participants were referred from primary care units, but some referrals came from other clinics at the hospital Sources of recruitment: memory clinic, Sweden
Index tests	CSF t-tau; CSF p-tau; CSF t-tau/ABeta ratio CSF was collected at baseline in polypropylene tubes and gently mixed to avoid gradient effects. All samples were centrifuged within 30 minutes at +4 uC at 2000 g for 10 min to remove cells and debris. Samples were stored in aliquots at -80 ^{o}C pending biochemical analysis. The procedure used and the analysis of the CSF followed the Alzheimer's Association Flow Chart for lumbar puncture (Blennow 2010) . The Luminex xMAP technology was used to determine the levels of tau, ABeta42 and p-tau (Ollson 2005). In addition to tau, ABeta42 and p-tau, the ratio of ABeta42/tau was tested as a separate variable in the logistic regression models since it previously had shown high predictive accuracy in this cohort (Hertze 2010) . Lumbar puncture was only conducted at the initial visit Threshold: CSF t-tau: > 87 pg/mL; CSF p-tau: > 39 pg/mL. The cut-offs were optimised (page e38639) Not reported whether the index test results were interpreted without knowledge of the results of the reference standard
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia or other forms of dementia Reference standards: NINCDS-ADRDA for AD; NINDS-AIREN/Erkinjuntti for VaD; McKeith for DLB Clinicians conducting follow-up were not aware of CSF biomarker results (page e38639)
Flow and timing	Duration of follow-up: mean 5.9 years (range: 3.2 to 8.8) At baseline: 133 MCI participants At follow-up: 62 MCI-MCI; 52 MCI-AD; 19 MCI-other forms of dementias (Table, 2; page e38639) Number included in analysis: 133 Conversion to Alzheimer's disease dementia: 1) CSF t-tau: sensitivity = 80%; specificity = 72% (Table, 2; page e38639) TP = 42; FP = 23; FN = 10; TN = 58 (calculated in RevMan5) 2) CSF p-tau: sensitivity = 67%; specificity = 86% (Table, 2; page e38639) TP = 35; FP = 11; FN = 17; TN = 70 (calculated in RevMan5)

	Note: the accuracy of the CSF t-tau/ABeta ratio not reported Loss to follow-up: none		
Comparative			
Notes	The author contacted regarding the sample procedure (Dr Palmqvist email on 28/2/14). Data for CSF t-tau/ABeta ratio not available		

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
		Unclear	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
		High	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**

Palmqvist 2012 (*Continued*)

DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Low	

Parnetti 2012

Study characteristics	
Patient sampling	Study design: nested case-control study with a delayed verification design 454 participants were consecutively referred to the memory clinic for a first diagnostic assessment of cognitive disturbances during the period 2005 to 2007. A prospective MCI group of 90 participants were recruited. 28 participants with AD were also enrolled in the study. We only included data on performance of the index test to discriminate between participants with MCI who converted to dementia and those who remained stable No exclusion criteria were specified.
Patient characteristics and setting	90 participants diagnosed by the Petersen 1999 criteria at baseline and had follow-up assessment at least once a year during four years Gender: total: 66 men; 34 women Age (y): MCI-MCI: mean 66.35 ± 8.22; MCI-AD: mean 67.23 ± 9.04 APOE $\epsilon 4$ carrier: not reported MMSE: MCI-MCI: 27.28 ± 1.47; MCI-AD: 26.66 ± 1.58 Sources of referral: not reported Sources of recruitment: memory clinic, Clinica Neurologica, Universita degli Studi di Perugia, Italy
Index tests	CSF p-tau/ABeta$_{1-42}$ ratio; CSF t-tau; CSF p-tau; CSF t-tau/ABeta$_{1-42}$ ratio Data available only for the CSF p-tau/ABeta$_{1-42}$ ratio. Authors contacted. Lumbar puncture was performed after an overnight fasting. CSF (10 mL) was collected in sterile polypropylene tubes, centrifuged for 10 min at 3000 × g and divided in 0.5 mL aliquots which were immediately frozen at−80 ^{o}C. CSF A1-40 was measured using a commercially available ELISA (IBL International, Japan) following instruction from the manufacturer. CSF ABeta$_{1-42}$, total tau, and p-tau were measured with ELISA method (Innotest ABeta$_{1-42}$, hTAU-Ag, p-tau 181 Ag, Innogenetics NV, Gent, Belgium) (Andreasen 1999; Blennow 1995) Threshold: 1074.0 for CSF p-tau/ABeta$_{1-42}$ ratio. Cutoff values were calculated using sensitivity and specificity values that maximized Youden's index (Youden 1950) Not reported whether the index test results were interpreted without knowledge of the results of the reference standard

Parnetti 2012 (*Continued*)

Target condition and reference standard(s)	Target condition: conversion from MCI to Alzheimer's disease Reference standard: not specified. MCI participants were clinically evaluated at least once a year during 4-year follow-up period (p 230). However, it was reported that the NINCDS-ADRDA criteria were used at baseline to identify AD diagnostic group
Flow and timing	Duration of follow-up (y): mean 3.40 ± 1.01 (maximum 4 years) At baseline: 90 MCI participants At follow-up: 90 MCI: 32 MCI-AD; 58 MCI-MCI (stable); (page 230) Number included in analysis: 90 Conversion to Alzheimer's disease dementia: CSF p-tau/ABeta1-42 ratio: sensitivity 81%; specificity 95% (page 233) TP = 26; FP = 3; FN = 6; TN = 55 (calculated in RevMan5) Loss to follow-up: none
Comparative	
Notes	Trial investigators contacted. Missing data requested for CSF t-tau, CSF p tau and CSF t-tau/ABeta1-42 ratio biomarkers. No further information was available at the time this review was prepared

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
		Unclear	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	No		
		High	**Low**

Parnetti 2012 (*Continued*)

DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	**Unclear**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Unclear		
Were all patients included in the analysis?	Yes		
		Unclear	

Visser 2009

Study characteristics	
Patient sampling	Study design: nested case-control study with a delayed verification design Prospective group of participants with SCI, naMCI and aMCI were recruited from 20 memory clinics across Europe, between January 2003 and June 2005, into the prospective DESCRIPA cohort study. Neurologically healthy controls were also recruited. Sampling procedure for a MCI cohort not described. We only included data on performance of the index test to discriminate between participants with MCI who converted to dementia and those who remained stable Exclusion criteria were a diagnosis of dementia or any somatic, psychiatric, or neurological disorder that might have caused the cognitive impairment
Patient characteristics and setting	168/193 participants from the DESCRIPA cohort with an available CSF baseline sample were included in the study. The data overlapped with the data from the Vos 2013 paper. However, we considered different CSF biomarkers in those two studies Gender: 88 men; 80 women Age (years): 70.0 ± 7.7 naMCI; 70.0 ± 7.7 aMCI; 66.0 ± 7.9 SCI MMSE: 27.6 ± 2.2 naMCI; 25.9 ± 2.8 aMCI; 28.8 ± 1.2 SCI Sources of recruitment: European multicentre memory clinics

Visser 2009 (*Continued*)

Index tests	CSF p-tau; CSF p-tau/ABeta ratio CSF was collected by lumbar puncture, centrifuged, and stored at -80 °C in polypropylene tubes, except where specified. The investigators measured CSF biomarkers with single-parameter ELISA kits (Innotest β-amyloid [1-42]; Innotest hTAU-Ag; Innotest Phospho-tau [181P]; Innogenetics, Ghent, Belgium). Analyses were done at one laboratory (Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Sweden) by operators who were blinded to all clinical information Threshold (positive test): CSF p-tau: i) ≥ 51 pg/mL (used in clinic) and ii) ≥ 85 pg/mL (> 90th percentile controls after correction for age); CSF p-tau/ABeta ratio: ≤ 9.92 (< 10th percentile of reference group after correction for age)
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia Reference standard: NINCDS-ADRDA criteria; DSM-IV criteria Diagnosis of dementia was conducted blinded to results of CSF biomarker analysis (page 621)
Flow and timing	Duration of follow-up: range 1 to 3 years 193 participants in the DESCRIPA cohort had CSF samples collected. Twenty-five participants were not included in the study: 6 had no neuropsychological test done at baseline; 11 had CSF collected at follow-up but not at baseline; 8 had insufficient CSF left for central analysis At baseline: N = 168 (MCI = 108 and SCI = 60) Number included in analysis: N = 158 Conversion to AD: 1.a) CSF p-tau threshold: ≥ 51 pg/mL (used in clinic). TP = 31; FP = 77; FN = 4; TN = 46 (unpublished data obtained from the author); sensitivity = 88%; specificity = 37% (calculated in Revman5) 1.b) CSF p-tau threshold: ≥ 85 pg/mL (> 90th percentile controls after correction for age) TP = 20; FP = 25; FN = 15; TN = 98 (unpublished data obtained from the author); sensitivity = 58%; specificity = 80% (calculated in Revman5) 2) CSF p-tau/ABeta ratio threshold: ≤ 9.92 (< 10th percentile of controls after correction for age) TP = 28; FP = 49; FN = 7; TN = 74 (unpublished data obtained from the author); sensitivity = 80%; specificity = 60% (calculated in Revman5) Loss to follow-up: 10 (CSF follow-up data were not available; the reason not given)
Comparative	
Notes	The trial investigators contacted; they provided requested data tor the 2 x 2 table to be completed; email from Dr Visser on 14/4/14

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		

Did the study avoid inappropriate exclusions?	Yes		
		Unclear	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
		Unclear	

Vos 2013

Study characteristics	
Patient sampling	Study design: prospective cohort study 399 participants with aMCI and 226 participants with naMCI from the DESCRIPA cohort and Alzheimer Centre of the VU University medical centre (VUmc). DESCRIPA is a European multicentre study performed in a memory clinic setting and enrolled subjects between 2003 and 2005. For this study, 431 eligible subjects were selected from 16 participating centres in which CSF was collected, MRI was performed, or APOE genotype was determined. The VUmc centre was one of the DESCRIPA centres and contributed an additional sample of 194 subjects that were seen outside the DESCRIPA inclusion period with data on CSF, MRI, or APOE measures. No differences in biomarkers were found between subjects from the VUmc centre in the DESCRIPA study and those from the additional VUmc sample. The sampling procedure was not described Exclusion criteria were diagnosis of dementia at baseline or any other somatic, psychiatric or neurological disorder that might have caused the cognitive impairment
Patient characteristics and setting	231/635 had available CSF data and were included in the review. 214/235 MCI participants, diagnosed by Petersen 2004 criteria at baseline, had a least one follow-up assessment. Baseline demographic data reported on all 625 participants. There was some data overlap with the data from the Visser 2009 paper. However, we considered different CSF biomarkers from those two studies Gender: 270 men; 335 women Age: 70.7 ± 7.6 years naMCI; 70.7 ± 7.8 aMCI MMSE: 27.5 ± 2.1 naMCI; 26.5 ± 2.5 aMCI Sources of recruitment: European multicentre memory clinics
Index tests	CSF t-tau; CSF t-tau/ABeta ratio CSF was collected by lumbar puncture, centrifuged, and stored at -80 °C in polypropylene tubes. Three samples were thawed twice but analyses without these samples revealed similar results. CSF ABeta1-42 and total tau (t-tau) were measured by experienced technicians using commercially available sandwich ELISAs (Innotest ABeta-amyloid 1-42; Innotest hTAU-Ag; Innogenetics, Ghent, Belgium), specially constructed to measure ABeta-amyloid 1-42 and t-tau, at the lab in Gothenburg for the DESCRIPA cohort and in Amsterdam for the additional subjects of the VUmc cohort. We corrected for inter-laboratory ELISA differences by analysing 33 samples at both labs and we adjusted VUmc values to those of DESCRIPA using the following formula: Gothenborg = (SD Gothenborg/SD VUmc) * VUmc + average Gothenborg - ((SD Gothenborg/SD VUmc) * average VUmc) Threshold (positive test): CSF t-tau: > 450 pg/mL for age less than 70 years; > 500 pg/mL for age older than 70 years; CSF t-tau/ABeta ratio: ABeta1-42/(240 1 [1.18 3 t-tau]) 1.0 Index test was conducted before follow-up.
Target condition and reference standard(s)	Target condition: Alzheimer's disease dementia Reference standard: NINCDS-ADRDA criteria; DSM-IV criteria Diagnosis of dementia was conducted blinded to results of CSF biomarker analysis (page 8)
Flow and timing	Duration of follow-up: mean 2.5 years (maximum duration 5 years); follow-up was performed annually At baseline: 231 MCI At follow-up: 214 MCI: 91 MCI-AD; 123 MCI-MCI Number included in analysis: 214

	Conversion to Alzheimer's disease dementia (data obtained from Dr Vos): 1) CSF t-tau TP = 65; FP = 28; FN = 26; TN = 95; sensitivity = 71%; specificity = 77% 2) CSF t-tau/ABeta ratio TP = 87; FP = 60; FN = 4; TN = 63; sensitivity = 96%; specificity = 51% Loss to follow-up: 17 participants did not have a follow-up assessment (some refused to participate or were untraceable or died before follow-up)
Comparative	
Notes	The trial investigators contacted; they provided requested data tor the 2 x 2 table to be completed; email from Dr Vos on 14/4/14

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
		Unclear	**Low**
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	**Low**
DOMAIN 3: Reference Standard			
Is the reference standards likely to correctly classify the target condition?	Yes		

Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
		Low	**Low**
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
		Unclear	

AD:*Alzhemer*′*sdisease*;*ADD*:*Alzheimer*′*sdiseasedementia*;*ADL*:*activitiesofdailyliving*;*aMCI*:*amnesticmildcognitiveimpairment*;*APOE*ϵ4:
Apolipoprotein E epsilon-4; CDR: clinical dementia rating; CNS: central nervous system; CSF: cerebrospinal fluid; DLB: Dementia with Lewy Bodies; DSM-III-R: Diagnostic and Statistical Manual of Mental Disorders-III-Revised; DSM-IV: Diagnostic and Statistical Manual of Mental Disorders-IV-Revised; ELISA: enzyme-linked immunoabsorbent assay; FTD: fronto-temporal dementia; GDS: Global Deterioration Scale; IQR: interquartile range; LP: lumbar puncture; MCI: mild cognitive impairment; MMSE: mini-mental state examination; nAD: non-Alzheimer's disease; naMCI: non-amnestic mild cognitive impairment; NINCDS-ADRDA: National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association; NINDS-AIREN: National Institute of Neurological Disorders and Stroke and Association and Internationale pour la Recherché et l'Enseignement en Neurosciences; ROC: receiver operating characteristics; SCI: subjective cognitive impairment; VaD: vascular dementia

Notes:Mattson 2009 is important and one of the single most defining studies in the field. However, we were not able to include it in our review because there was an overlap between participants in the Mattsson 2009 paper and participants assessed in the seven studies included in our review: Buchhave 2012; Eckerstrom 2010; Hansson 2006; Herukka 2007; Kester 2011; Palmqvist 2012; Visser 2009; therefore, we identified this paper as a 'multiple publication' and added it to those seven studies

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Desikan 2011	Not having data for constructing a 2 X 2 table. Study design: target condition was not conversion from MCI to dementia. The focus of the study was to asses relationship between neurodegeneration, amyloid Aß and CSF t-tau in MCI and healthy elderly controls. ADNI participants

(*Continued*)

Forlenza 2010	Not having data for constructing a 2 X 2 table. Study design: threshold not used. The objective of the study was to examine CSF biomarker levels between MCI-AD converters and MCI-MCI stable participants
Holland 2012	Not having data for constructing a 2 X 2 table. Study design: threshold not used. Annual changes in CSF biomarker levels were considered.The focus of the study was to asses the effects of age on rates of clinical decline
Ivanoiu 2005	Not having data for constructing a 2 X 2 table. Study design: threshold not used. The objective of the study was to examine prediction of progression to Alzheimer's disease and correlation with neuropsychological examination
Jack 2011	ADNI study. Not having data for constructing a 2 X 2 table. Study design: threshold not used. Target condition was not conversion from MCI to dementia. The objective of the study was to empirically assess the concept that Alzheimer's disease biomarkers significantly depart from normality in a temporarily ordered manner
Jagust 2009	ADNI study. Not having data for constructing a 2 X 2 table. Study design: threshold not used. Target condition was not conversion from MCI to dementia. The objective of the study was to assess relationship between biomarkers in ageing and dementia
Lanari 2009	Insufficient data to complete 2 X 2 tables. Additional data were requested from the trial investigators but no further information was available at the time this review was prepared
Landau 2010	ADNI study. Insufficient data to complete 2 X 2 tables. Additional data were requested from the trial investigators but no further information was available at the time this review was prepared
Maruyama 2004	Insufficient data to complete 2 X 2 tables. Additional data were requested from the trial investigators but no further information was available at the time this review was prepared
Maruyama 2004b	Insufficient data to complete 2 X 2 tables. Additional data were requested from the trial investigators but no further information was available at the time this review was prepared
Mattsson 2012	Not having data for constructing a 2 X 2 table. Study design: retrospective analysis. The objective of the study was to evaluate changes in biomarker levels between MCI-AD converters and MCI-MCI stable participants over time Participants: 15 MCI-AD and 15 MCI-MCI participants selected from a 4-year follow-up study
Nordlund 2010	Insufficient data to complete 2 X 2 tables. Additional data were requested from the trial investigators but no further information was available at the time this review was prepared
Okamura 2002	Insufficient data to complete 2 X 2 tables. Additional data were requested from the trial investigators but no further information was available at the time this review was prepared
Okonkwo 2011	ADNI study. Not having data for constructing a 2 X 2 table. Index test: combined CSF biomarkers. The relevant data for each individual CSF biomarker were not available
Pereira 2010	Not having data for constructing a 2 X 2 table. Study design: threshold not used. The objective of the study was to examine the pattern of functional impairment in the continuum MCI-AD

(*Continued*)

Perneczky 2011	Not having data for constructing a 2 X 2 table. Study design: threshold not used. The accuracy of CSF t-tau not evaluated. CSF t-tau levels measured in different diagnostic groups
Riemenschneider 2002	Insufficient data to complete 2 X 2 tables. Additional data were requested from the trial investigators but no further information was available at the time this review was prepared
Samtani 2012	ADNI study. Not having data for constructing a 2 X 2 table. Study design: threshold not used. Target condition was not conversion from MCI to dementia. The objective of the study was to develop a semi-mechanistic disease progression model for MCI participants
Schneider 2010	Not having data for constructing a 2 X 2 table. Study design: target condition was not conversion from MCI to dementia. The aim of the study was to test the recommendation of including MCI participants with low CSFamyloid ABeta and high CSF t-tau/ABeta ratio biomarkers in clinical trials, in order to improve efficiency of the RCT
Shaw 2009	ADNI study. Not having data for constructing a 2 X 2 table. The aim of the study was to develop a cerebrospinal fluid biomarker signature for mild Alzheimer's disease in ADNI participants
Sluimer 2010	Not having data for constructing a 2 X 2 table. Study design: threshold not used. Annual changes in CSF biomarker levels were considered. The focus of the study was to assess the association between CSF biomarker levels and MRI-based whole brain atrophy rate in MCI and AD
Snider 2009	Not having data for constructing a 2 X 2 table. Study design: threshold not used. The aim of the study was to determine whether baseline CSF biomarker's levels predict a rate of cognitive change, measured by CDR-SB (Clinical Dementia Rating Sum of Boxes score) in participants with very mild DAT (CDR = 0. 5 and Berg 1998 standard criteria)
Van Harten 2012	Not having data for constructing a 2 X 2 table. Study design: target condition was clinical progression, not conversion from MCI (SMC) to Alzheimer's disease dementia, in participants with cognitive complaints. According to our inclusion criteria, we considered participants with 'subjective memory complaints' (Matthew 2008)
Verwey 2008	Not having data for constructing a 2 X 2 table. Study design: threshold not used. Target condition not conversion from MCI to dementia. The focus of the study was to evaluate changes in CSF levels of tau and p-tau over time
Walhovd 2010	ADNI study. Not having data for constructing a 2 X 2 table. Study design: threshold not used. Target condition was not conversion from MCI to Alzheimer's disease dementia. The aim of the study was to assess an overall classification accuracy of biomarkers for diagnostic groups (for instance, controls vs AD), or to assess predictive accuracy of clinical change in MCI
Wang 2012	Not having data for constructing a 2 X 2 table. Study design: threshold not used. Target condition was not conversion from MCI to Alzheimer's disease dementia. The aim of the study was to determine whether CSF proteins were associated with hippocampal degeneration in participants with clinically diagnosed early AD

(*Continued*)

Westman 2012	ADNI study. Insufficient data to complete 2 X 2 tables. Additional data were requested from the trial investigators. The accuracy of the combination of the three CSF biomarkers (ABeta42, t-tau and p-tau) was assessed. The author could not provide us with the relevant data for each individual CSF biomarker
Yang 2012	ADNI study. Not having data for constructing a 2 X 2 table. The accuracy of the combined CSF biomarkers, as well as the accuracy of the combination of those CSF and structural biomarkers were assessed. The relevant data for each individual CSF biomarker were not reported

AD:Alzheimer'sdisease;ADNI:AlzheimerDiseaseNeuroimagingInitiative;CDR−SB:ClinicalDementiaRatingSumof Boxesscore;DAT:dementiaAlzheimer'stype;MCI:mildcognitiv

Characteristics of studies awaiting classification *[ordered by study ID]*

Balasa 2014

Study characteristics	
Patient sampling	Study design: nested case-control study with a delayed verification design 120 participants with early onset of cognitive impairment (51 MCI; 42 AD; 10 FTD; 3 posterior cortical atrophy; 14 primary progressive aphasia), who were referred to outpatient clinic at the Hospital Clinic Barcelona, were recruited prospectively between January 2009 and March 2013. Thirty-seven control subjects were also recruited. We only included data on performance of the index test to discriminate between participants with MCI who converted to dementia and those who remained stable Exclusion criteria: not reported.
Patient characteristics and setting	51 participants with MCI (25 amnestic; 20 amnestic multi-domain; 6 non-amnestic) diagnosed by Petersen 2004 criteria will be included in an updated review GENDER: 28 men; 23 women AGE (y): 57.9 ± 6 (range = 37 to 66) APOE 4 carrier (%): 37.5 MMSE (SD): 25.6 ± 6 Education: not reported Sources of referral: not reported Sources of recruitment: Alzheimer's Disease and Other Cognitive Disorders Unit outpatient clinic at the Hospital Clinic Barcelona, Italy
Index tests	CSF ABeta42/p-tau ratio All participants underwent lumbar puncture during the morning. The samples were centrifuged and stored in polypropylene tubes at −80 ºC within 2 hours. Levels of ABeta42, t-tau, and p-tau were measured by experienced laboratory personnel using commercial sandwich ELISA kits (Innogenetics, Gent, Belgium) Threshold: prespecified; the ABeta42/p-tau ratio was used in order to classify all the subjects as CSF positive (ratio ABeta42/p-tau < 6.43) or negative (ratio ABeta42/p-tau ≥ 6.43)

Balasa 2014 (*Continued*)

Target condition and reference standard(s)	Target condition: conversion from MCI to Alzheimer's disease dementia Reference standard: NIA-AA workgroup recommendations
Flow and timing	Duration of follow-up: 41 months for MCI-AD; 30 months for MCI-MCI At baseline: 51 MCI: 25 MCI with positive CSF biomarker; 26 MCI with negative CSF biomarker (Figure 1, p 924) At follow-up: 25 MCI with positive CSF biomarker: 24 MCI-AD and 1 MCI nonconverters; 26 MCI with negative CSF biomarker: 26 MCI nonconverters; 0 MCI converters; D^+ (disease positive) = 24; D^- (disease negative) = 27 Number to be included in analysis: (N = 51) Conversion to ADD: TP = 24; FP = 1; FN = 0; TN = 26 (Fig 1, p 924) sensitivity = 100%; specificity = 96% (calculated in RevMan5)
Comparative	
Notes	Authors need to be contacted in order to confirm that our calculation based on the information from Figure 1, p 924 is correct Participants: MCI participants with early onset of cognitive impairment (age < 65 years). This needs to be taken into consideration if the study is going to be included in an updated review

Eckerstrom 2015

Study characteristics	
Patient sampling	Study design: retrospective analysis of the longitudinal data; this is a sub-study of the Gothenborg MCI study We included data on performance of the index test to discriminate between participants with MCI who converted to dementia and those who remained stable Exclusion criteria: age > 79 or < 40, Mini Mental State Examination (MMSE) score < 19, acute/unstable somatic disease, severe psychiatric disorder, substance abuse or confusion caused by drugs
Patient characteristics and setting	73 participants diagnosed with the Global Deterioration Scale (GDS) criteria (Reisberg 1998) at baseline: 34 MCI converters (18 MCI-AD; 16 MCI-non-AD) and 39 MCI-stable GENDER: 14 men and 25 women MCI-stable; 13 men and 21 women MCI converters AGE (y): 64.4 ± 7.0 MCI-stable; 66.9 ± 6.9 MCI converters APOE 4 carrier (%): 16/39 (41%) MCI-stable; 24/34 (71%) MCI converters MMSE: 28.6 ± 1.4 MCI-stable; 27.6 ± 2.0 MCI converters Education: 12.4 ± 3.8 MCI-stable; 11.3 ± 4.0 MCI converters Sources of referral: not reported Sources of recruitment: University of Gothenburg, Molndal, Sweden
Index tests	CSF p-tau; CSF t-tau/ABeta42 ratio CSF samples were collected by lumbar puncture, which was performed in the morning to exclude influence on the results from possible diurnal fluctuations in biomarker levels. CSF samples were collected in a polypropylene tube, and immediately transported to the local laboratory for centrifugation. They were stored at −80 ºC, without being thawed and refrozen, pending biochemical analyses. CSF T-tau, P-tau181, and ABeta42 levels were determined using sandwich enzyme-

Eckerstrom 2015 (*Continued*)

	linked immunosorbent assays (INNOTEST® hTau Ag, INNOTEST® PHOSPHO-TAU(181P), and INNOTEST® !-AMYLOID(1-42), respectively) from Innogenetics Threshold: CSF p-tau: 73; CSF t-tau/ABeta42 ratio: 0.85 (p 207); not prespecified. ROC curves were used to calculate the cutoff values based on the maximum for the sum of sensitivity and specificity
Target condition and reference standard(s)	Target condition: conversion from MCI to Alzheimer's disease dementia or 'all dementia' Reference standard: NINCDS-ADRDA criteria
Flow and timing	Duration of follow-up: 43.1 ± 23 months MCI-stable; 33.7 ± 24 months MCI converters At baseline: 73 MCI participants At follow-up: 34 MCI converters (18 MCI-AD; 16 MCI-non-AD) and 39 MCI-stable Note: One patient (MCI-stable) declined LP. Additionally, analysis of CSF p-tau could not be performed on 9 patients (7 MCI converters, 2 MCI-stable) due to lack of CSF Table 2, p 208: CSF p-tau reported in 63/73 participants; number of MCI participants with the CSF t-tau/ABeta Beta42 ratio value not reported Number included in analysis: (N = 63) CSF p-tau: at follow-up, 27 MCI converters; 36 MCI-stable; D^+ (disease positive) = 27; D^- (disease negative) = 36 Conversion to 'all dementia': sensitivity = 44%; specificity = 92% (Table 3, p 209) TP = 12; FP = 3; FN = 15; TN = 33 (calculated in Revman5) Conversion to ADD: sensitivity = 75%; specificity = 92% (Table 4, p 209) Insufficient data to create 2 X 2 tables. It was not reported in which 7MCI (?MCI-AD; ?MCI-non-AD) CSF p-tau was not performed CSF t-tau/ABeta42 Note: The accuracy data of CSF t-tau/ABeta42 ratio biomarker not reported
Comparative	
Notes	Authors need to be contacted in order to obtain missing data. Check with the authors whether the sensitivity and specificity values given in Table 3 and Table 4 relate to a threshold given for CSF p-tau

Ewers 2012

Study characteristics	
Patient sampling	Study design: nested case-control study with a delayed verification design Subjects with a complete data set of CSF, MRI and neuropsychological tests were drawn from the ADNI data set including 130 participants with amnestic MCI, 81 participants with AD, and 101 elderly healthy controls. We only included data on performance of the index test to discriminate between participants with MCI who converted to dementia and those who remained stable Exclusion criteria: not reported.

Ewers 2012 (*Continued*)

Patient characteristics and setting	130 participants with amnestic MCI will be included in an updated review. Diagnostic criteria for amnestic MCI: subjective memory impairment and objective memory impairment identical to that for AD; a CDR of 0.5 including the memory box score of 0.5 or greater, and a MMSE score between 24 and 30; unimpaired general cognitive ability; functional performance such that participants with MCI did not meet criteria for dementia GENDER: 44 men and 28 women MCI-stable; 39 men and 19 women MCI-AD AGE (y): 73.4 ± 7.4 MCI-stable; 74.6 ± 7.3 MCI-AD APOE 4 carrier (%): 46 MCI-stable; 65.5 MCI-AD MMSE: 27.4 ± 1.6 MCI-stable; 26.9 ± 1.8 Education: not reported Sources of referral: not reported Sources of recruitment: ADNI participants
Index tests	CSF t-tau; CSF p-tau All CSF samples collected at the different centres were shipped on dry ice to the Penn ADNI Biomarker Core Laboratory at the University of Pennsylvania, Philadelphia for storage at -80 °C until further analysis at the laboratory. The concentration of CSF biomarkers was measured in the baseline CSF samples using Innogenetics reagents (research use only AlzBio3 immunoassay kits, Ghent, Belgium) and the multiplex xMAP Luminex platform (Lumnix Corporation, Austin, TX) at the Penn ADNI Biomarker Core Laboratory Threshold: not reported
Target condition and reference standard(s)	Target condition: conversion from MCI to Alzheimer's disease dementia Reference standard: NINCDS-ADRDA criteria
Flow and timing	Duration of follow-up: 2 years At baseline: 130 amnestic MCI At follow-up: 58 MCI-AD; 72 MCI-stable; D^+ (disease positive) = 58; D^- (disease negative) = 72 Number included in analysis: (N = 130) Conversion to ADD: CSF t-tau Sensitivity = 60.7%; specificity = 58.9% (Table 2, p 1209) TP = 35; FP = 30; FN = 23; TN = 42 (calculated in RevMan5) CSF p-tau Sensitivity = 63.9%; specificity = 58.9 (Table 2, p 1209) TP = 37; FP = 30; FN = 21; TN = 42 (calculated in RevMan5)
Comparative	
Notes	Authors need to be contacted in order to obtain threshold's values for CSF biomarkers; also check whether all 130 MCI participants were included in the analysis (Table 2, p 1029)

Leuzy 2015

Study characteristics	
Patient sampling	Study design: nested case-control study with a delayed verification design Thirty-three participants with MCI and thirty-five participants with AD were recruited from the Department of Geriatric Medicine, Karolinska University Hospital Huddinge, Stockholm, Sweden. We only included data on performance of the index test to discriminate between participants with MCI who converted to dementia and those who remained stable Exclusion criteria: not reported
Patient characteristics and setting	33 participants with MCI diagnosed by Petersen 1999 and Winbald 2004 criteria will be included in an updated review GENDER: 11 men and 10 women MCI-stable; 3 men and 9 women MCI-AD AGE (y): 63.52 ± 8.23 MCI-stable; 62.33 ± 6.96 MCI-AD APOE 4 carrier (%): not reported MMSE: not reported Education (y): 13.10 ± 3.24 MCI-stable; 13.58 ± 3.40 MCI-AD Sources of referral: not reported Sources of recruitment: Department of Geriatric Medicine, Karolinska University Hospital Huddinge, Stockholm, Sweden
Index tests	CSF t-tau; CSF p-tau; CSF t-tau/ABeta1-42 ratio; CSF p-tau/ABeta1-42 ratio CSF samples were obtained via lumbar puncture (LP) which was performed under non-fasting conditions, between 8 and 11 a.m., with a total of 10 mL of CSF collected. After discarding the first 0.5 mL, samples were centrifuged at 1500 × g (3000 to 4000 rpm) for 10 min at +4 ºC. Samples were then stored at -80 ºC in 1 mL portions pending biochemical analysis, without being thawed or refrozen. Levels of CSF biomarkers were determined using commercially available sandwich ELISAs (Innogenetics, Ghent, Belgium) Note: For the MCI group, levels of ABeta1-42 and t-tau were obtained for all subjects, with p-tau181p available for 26 of 33 subjects Threshold: prespecified; cutoff > 400 pg/mL CSF t-tau; cutoff > 80 pg/mL CSF p-tau; cutoff < 1.14 CSF ABeta1-42/t-tau ratio; cutoff < 6.5 CSF ABeta1-42/p-tau ratio
Target condition and reference standard(s)	Target condition: conversion from MCI to Alzheimer's disease Reference standard: NINCDS-ADRDA criteria
Flow and timing	Duration of follow-up: not reported At baseline: 33 MCI At follow-up: 21 MCI-stable; 12 MCI-AD Number included in analysis: (N = 33) Conversion to ADD: CSF t-tau TP = 8; FP = 7; FN = 4; TN = 14 (Table 2, p 1081) sensitivity = 67%; specificity = 67% (calculated in Revman5) CSF ABeta1-42/t-tau ratio TP = 6; FP = 6; FN = 6; TN = 15 (Table 2, p 1081) sensitivity = 50%; specificity = 71% (calculated in Revman5) Number included in analysis: (N = 26) Conversion to ADD: CSF p-tau

Leuzy 2015 (*Continued*)

	CSF ABeta1-42/p-tau ratio Insufficient data to create 2 X 2 tables. It was not reported how many MCI-stable and MCI-AD were at follow-up in a group of 26 MCI with available CSF p-tau biomarkers
Comparative	
Notes	Authors need to be contacted in order to obtain missing data for creating 2 X 2 tables for the CSF p-tau and CSF ABeta1-42/p-tau ratio biomarkers. Check with the authors whether the data used in 2 X 2 tables are correctly extracted from Table 2, p 1081 for the CSF t-tau and CSF ABeta1-42/t-tau ratio biomarkers; ask for a length of a follow-up period

AD:Alzheimer′sdisease;ADD:Alzheimer′sdiseasedementia;FTD:fronto−temporaldementia;GDS:GlobalDeteriorationScale;LP:lumbarpuncture;MCI:mildcognitiveimpairment;M

DATA

Presented below are all the data for all of the tests entered into the review.

Tests. Data tables by test

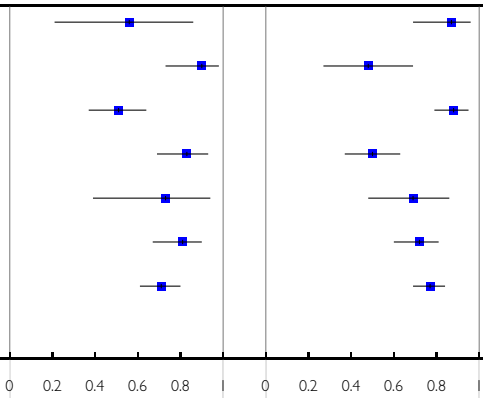
Test	No. of studies	No. of participants
1 CSF t-tau conversion to AD dementia	7	709
2 CSF p-tau conversion to AD dementia	6	492
3 CSF p-tau/ABeta ratio to AD dementia	5	433
4 CSF t-tau conversion to All dementias	4	319

Test 1. CSF t-tau conversion to AD dementia.

Review: CSF tau and the CSF tau/ABeta ratio for the diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI)

Test: 1 CSF t-tau conversion to AD dementia

Study	TP	FP	FN	TN	Sensitivity	Specificity
Amlien 2013	5	4	4	26	0.56 [0.21, 0.86]	0.87 [0.69, 0.96]
Hampel 2004	26	12	3	11	0.90 [0.73, 0.98]	0.48 [0.27, 0.69]
Hansson 2006	29	9	28	68	0.51 [0.37, 0.64]	0.88 [0.79, 0.95]
Kester 2011	35	29	7	29	0.83 [0.69, 0.93]	0.50 [0.37, 0.63]
Monge-Argiles 2011	8	8	3	18	0.73 [0.39, 0.94]	0.69 [0.48, 0.86]
Palmqvist 2012	42	23	10	58	0.81 [0.67, 0.90]	0.72 [0.60, 0.81]
Vos 2013	65	28	26	95	0.71 [0.61, 0.80]	0.77 [0.69, 0.84]

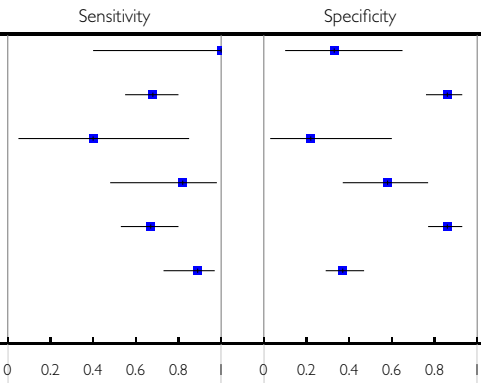


Test 2. CSF p-tau conversion to AD dementia.

Review: CSF tau and the CSF tau/ABeta ratio for the diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI)

Test: 2 CSF p-tau conversion to AD dementia

Study	TP	FP	FN	TN	Sensitivity	Specificity
Fellgiebel 2007	4	8	0	4	1.00 [0.40, 1.00]	0.33 [0.10, 0.65]
Hansson 2006	39	11	18	66	0.68 [0.55, 0.80]	0.86 [0.76, 0.93]
Koivunen 2008	2	7	3	2	0.40 [0.05, 0.85]	0.22 [0.03, 0.60]
Monge-Argiles 2011	9	11	2	15	0.82 [0.48, 0.98]	0.58 [0.37, 0.77]
Palmqvist 2012	35	11	17	70	0.67 [0.53, 0.80]	0.86 [0.77, 0.93]
Visser 2009	31	77	4	46	0.89 [0.73, 0.97]	0.37 [0.29, 0.47]

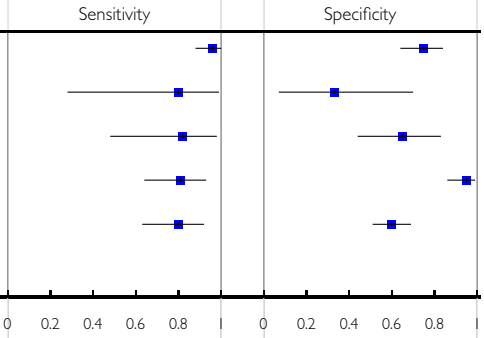


Test 3. CSF p-tau/ABeta ratio to AD dementia.

Review: CSF tau and the CSF tau/ABeta ratio for the diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI)

Test: 3 CSF p-tau/ABeta ratio to AD dementia

Study	TP	FP	FN	TN	Sensitivity	Specificity
Hansson 2006	55	19	2	58	0.96 [0.88, 1.00]	0.75 [0.64, 0.84]
Koivunen 2008	4	6	1	3	0.80 [0.28, 0.99]	0.33 [0.07, 0.70]
Monge-Argiles 2011	9	9	2	17	0.82 [0.48, 0.98]	0.65 [0.44, 0.83]
Parnetti 2012	26	3	6	55	0.81 [0.64, 0.93]	0.95 [0.86, 0.99]
Visser 2009	28	49	7	74	0.80 [0.63, 0.92]	0.60 [0.51, 0.69]

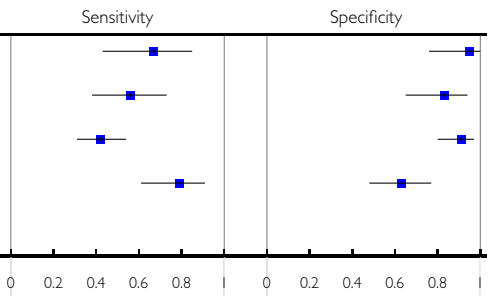


Test 4. CSF t-tau conversion to All dementias.

Review: CSF tau and the CSF tau/ABeta ratio for the diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI)

Test: 4 CSF t-tau conversion to All dementias

Study	TP	FP	FN	TN	Sensitivity	Specificity
Eckerstrom 2010	14	1	7	20	0.67 [0.43, 0.85]	0.95 [0.76, 1.00]
Galluzzi 2010	19	5	15	25	0.56 [0.38, 0.73]	0.83 [0.65, 0.94]
Hansson 2006	33	5	45	51	0.42 [0.31, 0.54]	0.91 [0.80, 0.97]
Herukka 2007	26	17	7	29	0.79 [0.61, 0.91]	0.63 [0.48, 0.77]



ADDITIONAL TABLES

Table 1. Studies awaiting classification

Conversion from MCI to Alzheimer's disease dementia							
Study	**Participants n/N (included in analysis)**	**Index test (number and % of positive tests)**	**Threshold (test abnormal) (prespecified Yes/No)**	**Number of converters (%) FP and FN**	**Test accuracy at study level**		**Duration of follow-up**
					Sensitivity	Specificity	
***Balasa 2014**	51/51	CSF ABeta42/p-tau ratio 25/51 (49%)	< 6.43 (Yes)	24/51 (47%) FP =1; FN =0	100%	96%	41 months for MCI-AD; 30 months for MCI-MCI
***Ewers 2012**	130/130	CSF t-tau 65/130 (50%)	Not reported	58/130 (45%) FP = 30; FN = 23	60.7%	58.9%	24 months
		CSF p-tau 67/130 (51.5%)	Not reported	58/130 (45%) FP = 30; FN = 21	63.9%	58.9%	
***Leuzy 2015**	33/33	CSF t-tau 15/33 (45%)	> 400 pg/mL (Yes)	12/33 (36%) FP = 7; FN = 4	67%	67%	Not reported
		CSF t-tau/ABeta ratio 12/33 (36%)	< 1.14 (Yes)	12/33 (36%) FP = 6; FN = 6	50%	71%	

Table 1. Studies awaiting classification (*Continued*)

Conversion from MCI to all dementias							
***Eckerstrom 2015**	73/73	CSF p-tau 15/73 (20.5%)	73 pg/mL (No)	27/73 (36.9%) FP = 3; FN = 15	75%	92%	43.1 ± 23 months MCI-stable; 33.7 ± 24 months MCI converters
Study awaiting translation							
Urakami 2004							

AD: Alzheimer's disease; FN: false negative; FP: false positive; MCI: mild cognitive impairment

∗Authors need to be contacted in order to obtain missing data / relevant information. Data presented are provisional.

Table 2. Conversion from MCI to Alzheimer's disease dementia

Included studies, index test and test accuracy at study level for conversion from MCI to Alzheimer's disease dementia							
Study	**Participants n/N (included in analysis)**	**Index test (number and % of positive tests)**	**Threshold (test abnormal) (prespecified Yes/No)**	**Number of converters (%) FP and FN**	**Test accuracy at study level**		**Duration of follow-up**
					Sensitivity	Specificity	
Amlien 2013	49/39	CSF t-tau 9/39 (23%)	≥ 300 ng/L for age younger than 50 years; ≥ 450 ng/L for age 50 to 69 years; ≥ 500 ng/L for age older than 70 years (Sjogren 2001) (Yes)	9/39 (23%); FP = 4; FN = 4	56%	87%	mean 2.6 ± 0.5 years (range 1.6 to 4 years)
Buchhave 2012*	137/134	CSF p-tau/ABeta ratio 69/134 (51%)	6.2 ng/L (No)	72/134 (54%) FP = 6; FN = 9	88%	90%	median: 9.2 years (range 4 to 12 years)
Fellgiebel 2007	16/16	CSF p-tau 12/16 (75%)	≥ 50 pg/mL (No)	4/16 (25%) FP = 8; FN = 0	100%	33%	mean 19.6 ± 9.0 months

Table 2. Conversion from MCI to Alzheimer's disease dementia (*Continued*)

Hampel 2004	52/52	CSF t-tau 38/52 (73%)	≥ 479 ng/L (No)	29/52 (56%); FP = 12; FN = 3	90%	48%	mean 8.4 ± 5.1 months (range 2 to 24 months)
Hansson 2006*	137/134	CSF t-tau 38/134 (28%)	> 350 ng/L (No)	57/134 (42%) ; FP = 9; FN = 28	51%	88%	Total sample: median 5. 2 years (range 4.0 to 6.8 years); MCI-AD: median: 4.3 years (range 1.1 to 6.7 years) MCI-other dementias: median 4.2 years (range 1.5 to 3 years)
		CSF p-tau 50/134 (37%)	≥ 60 ng/L (No)	57/134 (42%) ; FP = 11; FN = 18	68%	86%	
		CSF p-tau/ ABeta ratio 74/134 (55%)	6.5 ng/L (No)	57/134 (42%) ; FP = 19; FN = 2	96%	75%	
Kester 2011	153/100	CSF t-tau 64/100 (64%)	> 356 pg/mL (Yes)	42/100 (42%) FP = 29; FN = 7	83%	50%	median 18 months (IQR 13 - 24)
Koivunen 2008	15/14	CSF p-tau 9/14 (64%)	≥ 70 pg/mL (Yes)	5/14 (36%) FP = 7; FN = 3	40%	22%	2 years
		CSF p-tau/ ABeta ratio 9/14 (64%)	6.5 pg/mL (yes)	5/14 (36%) FP = 6; FN = 1	80%	33%	
Monge-Argiles 2011	37/37	CSF t-tau 16/37 (43%)	≥ 77.5 pg/mL (No)	11/37 (28%) FP = 8; FN = 3	73%	69%	6 months
		CSF p-tau 20/37 (54%)	≥ 54.5 pg/mL (No)	11/37 (28%) FP = 11; FN = 2	82%	58%	
		CSF p-tau/ ABeta ratio 18/37 (49%)	0.17 (No)	11/37 (28%) FP = 9; FN = 2	82%	66%	
		CSF t-tau/ ABeta ratio 23/37 (62%)	0.18 (No)	11/37 (28%) FP = 13; FN = 1	91%	50%	
Palmqvist 2013	133/133	CSF t-tau 65/133 (49%)	> 87 pg/mL (No)	52/133 (39%) FP = 23; FN = 10	81%	72%	mean 5.9 years (range 3.2 to 8.8 years)

Table 2. Conversion from MCI to Alzheimer's disease dementia (*Continued*)

		CSF p-tau 46/133 (34%)	> 39 pg/mL (No)	52/133 (39%) FP = 11; FN = 17	67%	86%	
Parnetti 2012	90/90	CSF p-tau/ ABeta ratio 29/90 (32%)	1074.0 (No)	32/90 (35%) FP = 3; FN = 6	81%	95%	maximum: 4 years; mean 3.40 ± 1.01 years
Visser 2009	168/158	CSF p-tau 108/158 (68%)	≥ 51 pg/mL (used in clinical practice) (No)	35/158 (22%) FP = 77; FN = 4	88%	37%	range 1 to 3 for MCI
		CSF p-tau 45/158 (28%)	≥ 85pg/mL (> 90th percentile of controls after correction for age) (No)	35/158 (22%) FP = 25; FN = 15	57%	80%	
		CSF p-tau/ ABeta ratio 77/158 (49%)	9.92 (< 10th percentile of reference group after correction for age) (No)	35/158 (22%) ; FP = 49; FN = 7	80%	60%	
Vos 2013	231/214	CSF t-tau 93/214 (43%)	> 450 pg/mL for age less than 70 years; > 500 pg/mL for age older than 70 years (Yes)	91/214 (42%) FP = 28; FN = 26	71%	77%	mean 2.5 ± 1.0 years
		CSF t-tau/ ABeta ratio 147/214 (69%)	ABeta1-42/ (240 1 [1.18 3 t-tau]) 1.0 (Yes)	91/214 (42%) FP = 60; FN = 4	96%	51%	

AD:Alzheimer'sdisease;FN:falsenegative;FP:falsepositive;MCI:mildcognitiveimpairment

∗*Studiesinvolvedthesameparticipants.OnlyHansson*2006*isincludedinthemeta−analysis*

Table 3. Conversion from MCI to All dementia

Included studies, index test and test accuracy at study level for conversion from MCI to All dementias							
Study	**Participants n/N (included in analysis)**	**Index test (Number and % of positive tests)**	**Threshold (test abnormal) (prespecified Yes / No)**	**Number of converters (%) FP and FN**	**Test accuracy at study level**		**Duration of follow-up**
					Sensitivity	Specificity	
Eckerstrom 2010	42/42	CSF t-tau 15/42 (36%)	≥ 500 ng/L (No)	21/42 (50%) FP = 1 FN = 7	67%	95%	Total sample: 19.6 ± 9.0 months; MCI-MCI: 19.5 ± 9.3 months; MCI-progressive: 17.6 ± 8.8 months (4/8 MCI-AD: 23.7 ± 2.0 months)
Galluzzi 2010	90/64	CSF t-tau 24/64 (37.5%)	> 450 pg/mL for subjects with an age range between 51 and 70 determined; > 500 pg/mL for subjects with an age range between 71 and 93 (Yes)	34/64 (53%) FP = 5 FN = 15	56%	83%	Total sample: 8.4 ± 5.1 months (range 2 to 24 months); follow-up interval for converters was 9.6 ± 5.4, and for non-converters 7.0 ± 4.3 months
Hansson 2006	137/134	CSF t-tau 38/134 (28%)	> 350 pg/mL (No)	78/134 (58%) FP = 5 FN = 45	42%	91%	Total sample: median 5.2 years (range 4.0 to 6.8); MCI-AD: median: 4.3 years (range 1.1 to 6.7); MCI-other dementias: median 4.2 (1.5 to 6.3)
Herukka 2007	79/79	CSF t-tau 43/79 (54%)	> 400 pg/mL (Yes)	33/79 (42%) FP = 17 FN = 7	79%	63%	Mean 3.52 ± 1.95 years in MCI converters; mean 4.56 ± 3.09 years in MCI-stable

APPENDICES

Appendix 1. Sources searched and search strategies

The MEDLINE search strategy below was created to optimise sensitivity. The strategy utilises a number of concepts:
Concept A: lines 1 to 21 health condition/s of interest
Concept B: lines 23 to 42 what is being measured by the index test/s/the index test/s
Concept C: lines 44 to 49 method of measurement (i.e. CSF)
The main yield is created by combining A AND B AND C
However, in order to try to capture those records that perhaps do not mention one or more of the three concepts above, some additional combinations were added to the strategy. For example: In the MEDLINE strategy below, lines 51 and 52 (which identify records in Medline with the dementia MeSH subheading of diagnosis and those with a subheading of cerebrospinal fluid) were combined with the concept for the index test(s). This approach identified unique records and an examination of the first 50 of these records resulted in two further citations for possible inclusion within the review.

Source	Search strategy	Hits retrieved
1. MEDLINE In-process and other non-indexed citations and MEDLINE 1946 to present (Ovid SP)	1. exp Dementia/ 2. Cognition Disorders/ 3. (alzheimer* or dement* or AD or lewy* or VaD or frontotemporal or 'vascular cognit* impair*').ti,ab 4. ((cognit* or memory or cerebr* or mental*) adj3 (declin* or impair* or los* or deteriorat* or degenerat* or complain* or disturb* or disorder*)).ti,ab 5. (forgetful* or confused or confusion).ti,ab. 6. MCI.ti,ab. 7. ACMI.ti,ab. 8. ARCD.ti,ab. 9. SMC.ti,ab. 10. CIND.ti,ab. 11. BSF.ti,ab. 12. AAMI.ti,ab. 13. LCD.ti,ab. 14. QD.ti,ab. 15. AACD.ti,ab. 16. MNCD.ti,ab. 17. MCD.ti,ab. 18. (nMCI or aMCI or mMCI).ti,ab. 19. ('N-MCI' or 'A-MCI' or 'M-MCI').ti,ab.	July 2012: 7718 Jan 2013: 480

(*Continued*)

	20. 'Petersen".ab. 21. ((CDR adj2 '0.5') or ('clinical dementia rating' adj3 '0.5')).ab 22. or/1-21 23. (neurofibril* adj3 tangle*).ti,ab. 24. (neurofilament adj3 protein*).ti,ab. 25. (neuropil adj3 thread*).ti,ab. 26. ((senile or amyloid or neuritic) adj3 plaque*).ti,ab. 27. Neuropil Threads/ 28. Senile Plaques/ 29. exp Neurofibrils/ 30. Neurofilament Proteins/ 31. tau Proteins/ 32. tau*.ti,ab. 33. hyperphosphorylation.ti,ab. 34. pTau181.ti,ab. 35. *peptide fragments/cf 36. pTau*.ti,ab. 37. ('t-tau*' or 'p-tau*').ti,ab. 38. (innotest or inno-bia or Alzbio3).ti,ab. 39. ((abeta* or ab42 or ab40 or 'amyloid-beta' or 'beta-amyloid' or 'a?42' or 'a?40' or 'a beta') adj4 (ratio or ratios)).ti,ab 40. ('phospho-tau*' or 'total-tau*').ti,ab. 41. or/23-40 42. (cerebrospinal fluid* or csf or 'spinal fluid*').ti,ab. 43. (blood or plasma).ti,ab. 44. Cerebrospinal Fluid/ 45. Blood-Brain Barrier/ 46. or/42-45 47. (cf or bl or di or du).fs. 48. or/46-47 49. 48 and 41 and 22 50. exp *Dementia/cf [Cerebrospinal Fluid] 51. exp Dementia/di [Diagnosis] 52. cf.fs. 53. 41 and 51 and 52 54. Cerebrospinal Fluid Proteins/ 55. Biological Markers/cf [Cerebrospinal Fluid] 56. or/54,55 57. 56 and 22 and 41 58. or/49,50,53,57 59. (animals not (humans and animals)).sh. 60. 58 not 59	

(*Continued*)

2. Embase 1980 to 2012 week 29 (Ovid SP)	1. dement*.ti. 2. alzheimer*.ti. 3. (AD or VaD or lewy or frontotemporal or 'vascular cognit* impair*').ti 4. Dementia/di 5. dementia/ep [Epidemiology] 6. (('conversion to' or 'conversion from') adj4 (dement* or alzheimer* or AD or lewy or VaD or 'vascular cognit* impair*')).ab 7. ((endpoint* or 'end point*' or outcome*) adj5 (dement* or alzheimer* or AD or VaD or lewy)).ab 8. (predict* adj5 (dement* or alzheimer* or AD or VaD or lewy or 'vascular cognit* impair*')).ab 9. ((convert or converted) adj4 (dement* or alzheimer* or AD or lewy or VaD or 'vascular cognit* impair*')).ab 10. (progress* adj5 (dement* or alzheimer* or AD or VaD or lewy or 'vascular cognit* impair*')).ab 11. or/1-10 12. exp dementia/ 13. (alzheimer* or dement* or AD or lewy* or VaD or frontotemporal or 'vascular cognit* impair*').ti,ab 14. ((cognit* or memory or cerebr* or mental*) adj3 (declin* or impair* or los* or deteriorat* or degenerat* or complain* or disturb* or disorder*)).ti,ab 15. (forgetful* or confused or confusion).ti,ab. 16. MCI.ti,ab. 17. ACMI.ti,ab. 18. ARCD.ti,ab. 19. SMC.ti,ab. 20. CIND.ti,ab. 21. BSF.ti,ab. 22. AAMI.ti,ab. 23. LCD.ti,ab. 24. QD.ti,ab. 25. AACD.ti,ab. 26. MNCD.ti,ab. 27. MCD.ti,ab. 28. (nMCI or aMCI or mMCI).ti,ab. 29. ('N-MCI' or 'A-MCI' or 'M-MCI').ti,ab.	July 2012: 3692 Jan 2013: 732

(*Continued*)

	30. 'Petersen criteria'.ab. 31. ((CDR adj2 '0.5') or ('clinical dementia rating' adj3 '0.5')).ab 32. (neurofibril* adj3 tangle*).ti,ab. 33. (neurofilament adj3 protein*).ti,ab. 34. (neuropil adj3 thread*).ti,ab. 35. ((senile or amyloid or neuritic) adj3 plaque*).ti,ab. 36. neuropil thread/ 37. senile plaque/ 38. neurofilament/ 39. neurofilament protein/ 40. or/12-39 41. (cerebrospinal fluid* or csf or 'spinal fluid*').ti,ab. 42. (blood or plasma).ti,ab. 43. cerebrospinal fluid/ 44. blood brain barrier/ 45. or/41-44 46. tau protein/ 47. tau.ti,ab. 48. hyperphosphorylation.ti,ab. 49. pTau181.ti,ab. 50. tau181.ti,ab. 51. peptide fragment/ 52. ('abeta*/tau' and ratio).ab. 53. pTau*.ti,ab. 54. ('t-tau*' or 'p-tau*').ti,ab. 55. (innotest or inno-bia or Alzbio3).ti,ab. 56. ((abeta* or ab42 or ab40 or 'amyloid-beta' or 'beta-amyloid' or (amyloid and 'β') or 'aβ' or 'aβ42' or 'aβ40' or 'a beta') adj4 ratio).ti,ab 57. ('phospho-tau*' or 'total-tau*').ti,ab. 58. tau231.ti,ab. 59. or/46-58 60. 40 and 45 and 59 61. sensitivit*.ab. 62. specificit*.ab. 63. (ROC or 'receiver operat*').ab. 64. area under the curve/ 65. ('Area under curve' or AUC).ab. 66. (detect* adj3 (dement* or AD or alzheimer*)).ti,ab. 67. sROC.ab. 68. accura*.ti,ab. 69. (likelihood adj3 (ratio* or function*)).ab.	

(*Continued*)

	70. (conver* adj3 (dement* or AD or alzheimer*)).ti,ab. 71. ((true or false) adj3 (positive* or negative*)).ab. 72. ((positive* or negative* or false or true) adj3 rate*).ti,ab 73. reproducibility/ 74. diagnos*.ti. 75. diagnostic accuracy/ 76. or/61-75 77. 11 and 59 and 76 78. 60 or 77	
3. PsycINFO 1806 to July week 1 2012 (Ovid SP)	1. dement*.ti,ab. 2. alzheimer*.ti,ab. 3. (AD or VaD or lewy or frontotemporal or 'vascular cognit* impair*').ti,ab 4. exp Dementia/ 5. (('conversion to' or 'conversion from') adj4 (dement* or alzheimer* or AD or lewy or VaD or 'vascular cognit* impair*')).ab 6. ((endpoint* or 'end point*' or outcome*) adj5 (dement* or alzheimer* or AD or VaD or lewy)).ab 7. (predict* adj5 (dement* or alzheimer* or AD or VaD or lewy or 'vascular cognit* impair*')).ab 8. ((convert or converted) adj4 (dement* or alzheimer* or AD or lewy or VaD or 'vascular cognit* impair*')).ab 9. (progress* adj5 (dement* or alzheimer* or AD or VaD or lewy or 'vascular cognit* impair*')).ab 10. or/1-9 11. Prediction/ or Diagnosis/ 12. (dement* or alzheimer* or AD or VaD or lewy or frontotemporal or 'vascular cognit* impair*').ab 13. exp *Dementia/ 14. or/11-13 15. 10 or 14 16. exp Dementia/ 17. exp Cognitive Impairment/ 18. (alzheimer* or dement* or AD or lewy* or VaD or frontotemporal or 'vascular cognit* impair*').ti,ab 19. (forgetful* or confused or confusion).ti,ab. 20. MCI.ti,ab.	July 2012: 2645 Jan 2013: 464

(*Continued*)

	21. ACMI.ti,ab. 22. ARCD.ti,ab. 23. SMC.ti,ab. 24. CIND.ti,ab. 25. BSF.ti,ab. 26. AAMI.ti,ab. 27. LCD.ti,ab. 28. QD.ti,ab. 29. AACD.ti,ab. 30. MNCD.ti,ab. 31. MCD.ti,ab. 32. (nMCI or aMCI or mMCI).ti,ab. 33. ('N-MCI' or 'A-MCI' or 'M-MCI').ti,ab. 34. 'Petersen criteria'.ab. 35. ((CDR adj2 '0.5') or ('clinical dementia rating' adj3 '0.5')).ab 36. (neurofibril* adj3 tangle*).ti,ab. 37. (neurofilament adj3 protein*).ti,ab. 38. (neuropil adj3 thread*).ti,ab. 39. ((senile or amyloid or neuritic) adj3 plaque*).ti,ab. 40. exp Neurofibrillary Tangles/ 41. exp Senile Plaques/ 42. or/16-41 43. (cerebrospinal fluid* or csf or 'spinal fluid*').ti,ab. 44. (blood or plasma).ti,ab. 45. exp Cerebrospinal Fluid/ 46. exp Blood Brain Barrier/ 47. or/43-46 48. tau.ti,ab. 49. hyperphosphorylation.ti,ab. 50. pTau181.ti,ab. 51. tau181.ti,ab. 52. ('abeta*/tau' and ratio).ab. 53. pTau*.ti,ab. 54. ('t-tau*' or 'p-tau*').ti,ab. 55. (innotest or inno-bia or Alzbio3).ti,ab. 56. ((abeta* or ab42 or ab40 or 'amyloid-beta' or 'beta-amyloid' or (amyloid and 'β') or 'aβ' or 'aβ42' or 'aβ40' or 'a beta') adj4 ratio).ti,ab 57. ('phospho-tau*' or 'total-tau*').ti,ab. 58. tau231.ti,ab. 59. or/48-58 60. 42 and 59 61. 47 and 60	

(*Continued*)

	62. 15 and 59 63. 61 or 62	
4. BIOSIS Previews (Thomson Reuters Web of Science)	Topic = (tau OR p-tau OR t-tau OR pTau OR tTau OR hyperphosphorylation OR pTau181 OR 'phospho-tau*' or 'total-tau*' OR tau231) AND Topic = (dement* OR alzheimer* OR MCI OR 'cognit* impair*' OR 'CDR 0.5' OR 'petersen criteria' OR aMCI OR nMCI OR mMCI) AND Topic = (diagnosis OR sensitiv* OR specificit* OR ROC OR 'receiver operat*' OR 'Area under curve' or AUC OR sROC OR accura* OR 'follow*-up' OR 'positive predictive value*' OR 'negative predictive value*' OR longitudinal OR longitudinally) Timespan = All Years. Databases = BIOSIS Previews. Lemmatization = On	July 2012: 1775 Jan 2013: 206
5. Web of Science Core Collection, including Conference Proceedings Citation Index (Thomson Reuters Web of Science) (1945-present)	Topic = (tau OR p-tau OR t-tau OR pTau OR tTau OR hyperphosphorylation OR pTau181 OR 'phospho-tau*' or 'total-tau*' OR tau231) AND Topic = (dement* OR alzheimer* OR MCI OR 'cognit* impair*' OR 'CDR 0.5' OR 'petersen criteria' OR aMCI OR nMCI OR mMCI) AND Topic = (diagnosis OR sensitiv* OR specificit* OR ROC OR 'receiver operat*' OR 'Area under curve' or AUC OR sROC OR accura* OR 'follow*-up' OR 'positive predictive value*' OR 'negative predictive value*' OR longitudinal OR longitudinally) Timespan = All Years. Databases = Web of Science Core Collection Lemmatization = On	July 2012: 2205 Jan 2013: 234
6. LILACS (BIREME)	Hiperfosforilación OR hyperphosphorylation OR tau OR fosfo-tau OR phosphor-tau OR p-tau OR pTau181 OR tau181 OR tau231	July 2012: 126 Jan 2013: 3
7. CINAHL (EBSCOhost)	S1 TX dement* S2 TX AD OR VaD OR lewy OR frontotemporal OR 'vascular cognit* impair*' S3 TX alzheimer* S4 (MH 'Dementia/DI') S5 (MH 'Dementia/ET') S6 TX 'conversion to' N2 dement* S7 TX ('conversion from') N4 (dement*	July 2012: 591 Jan 2013: 59

(*Continued*)

	or alzheimer* or AD or lewy or VaD or 'vascular cognit* impair*') S8 TX (endpoint* or 'end point*' or outcome*) N5 (dement* or alzheimer* or AD or VaD or lewy) S9 TX predict* N5 (dement* or alzheimer* or AD or VaD or lewy or 'vascular cognit* impair*') S10 TX (convert or converted) N4 (dement* or alzheimer* or AD or lewy or VaD or 'vascular cognit* impair*') S11 TX progress* N5 (dement* or alzheimer* or AD or VaD or lewy or 'vascular cognit* impair*') S12 S1 or S2 or S3 or S4 or S5 or S6 or S7 or S8 or S9 or S10 or S11 S13 (MH 'Predictive Value of Tests') S14 TX dement* or alzheimer* or AD or VaD or lewy or frontotemporal or 'vascular cognit* impair*' S15 (MM 'Dementia+') S16 S14 or S15 S17 S13 and S16 S18 S12 or S17 S19 (MH 'Dementia+') S20 (MH 'Cognition Disorders') S21 TX alzheimer* or dement* or AD or lewy* or VaD or frontotemporal or 'vascular cognit* impair*' S22 TX (cognit* or memory or cerebr* or mental*) N3 (declin* or impair* or los* or deteriorat* or degenerat* or complain* or disturb* or disorder*) S23 TX forgetful* or confused or confusion S24 TX MCI S25 TX ACMI S26 TX ARCD S27 TX SMC S28 TX CIND S29 TX LCD S30 TX AACD S31 TX MNCD S32 TX MCD S33 TX nMCI or aMCI or mMCI S34 TX 'N-MCI' or 'A-MCI' or 'M-MCI' S35 TX 'Petersen criteria' S36 TX CDR N2 '0.5' S37 TX 'clinical dementia rating' N3 '0.5'	

(*Continued*)

	S38 TX neurofibril* N3 tangle* S39 TX neurofilament N3 protein* S40 TX neuropil N3 thread* S41 TX (senile or amyloid or neuritic) N3 plaque* S42 S19 or S20 or S21 or S22 or S23 or S24 or S25 or S26 or S27 or S28 or S29 or S30 or S31 or S32 or S33 or S34 or S35 or S36 or S37 or S38 or S39 or S40 or S41 S43 TX cerebrospinal fluid* or csf or 'spinal fluid*' S44 (MH 'Cerebrospinal Fluid') S45 S43 or S44 S46 TX tau S47 TX hyperphosphorylation S48 TX pTau181 S49 TX tau181 S50 TX (abeta* N3 tau) N4 ratio S51 TX (amyloid* N3 tau) N4 ratio S52 TX (ab42 N3 tau) N4 ratio S53 TX (ab40 N3 tau) N4 ratio S54 TX ('a beta' N3 tau) N4 ratio S55 TX pTau* S56 TX 't-tau*' or 'p-tau*' S57 TX ('aβ40' or 'a beta') N4 ratio S58 TX 'phospho-tau*' or 'total-tau*' S59 TX tau231 S60 S46 or S47 or S48 or S49 or S50 or S51 or S52 or S53 or S54 or S55 or S56 or S57 or S58 or S59 S61 S42 and S60 S62 S12 and S60 S63 S17 and S45 S64 S61 or S62 or S63	
TOTAL before de-duplication and first assessment		July 2012: 18752 Jan 2013: 1694

Appendix 2. Cross classification of test results and disease status (2X2)

Table 1: Conversion from MCI to Alzheimer's disease dementia

Index test information	Reference standard information	
	ADD present	**ADD absent**
Index test positive	Index test + who convert to ADD (TP)	Index test + who remain MCI (FP) & Index test + who convert to non-ADD (FP)
Index test negative	Index test - who convert to ADD (FN)	Index test - who remain MCI (TN) & Index test - who convert to non-ADD (TN)

Table 2: Conversion from MCI to non-Alzheimer's disease dementia

Index test information	Reference standard information	
	Non-ADD present	**Non-ADD absent**
Index test positive	Index test + who convert to non-ADD (TP)	Index test + who remain MCI (FP) & Index test + who convert to ADD (FP)
Index test negative	Index test - who convert to non-ADD (FN)	Index test - who remain MCI (TN) & Index test - who convert to ADD (TN)

Table 3: Conversion from MCI to any form of dementia

Index test information	Reference standard information	
	Any forms of dementia present	**Dementia absent**
Index test positive	Index test + who convert to any form of dementia (TP)	Index test + who remain MCI (FP)
Index test negative	Index test - who convert to any form of dementia (FN)	Index test - who remain MCI (TN)

Appendix 3. Assessment of methodological quality table QUADAS-2 tool

DOMAIN	PARTICIPANT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING
Description	Describe methods of participant selection: Describe included participants (prior testing, presentation, intended use of index test, and setting)	Describe the index test and how it was conducted and interpreted	Describe the reference standard and how it was conducted and interpreted	Describe any participants who did not receive the index test(s) and/or reference standard or who were excluded from the 2 x 2 table (refer to flow diagram): Describe the time interval and any interventions between index test(s) and reference standard
Signalling questions: (yes/no/unclear)	Was a consecutive or random sample of participants enrolled?	Were the index test results interpreted without knowledge of the results of the reference standard?	Is the reference standard likely to correctly classify the target condition?	Was there an appropriate interval between index test(s) and reference standard?
	Was a case-control design avoided?	If a threshold was used, was it prespecified?	Were the reference standard results interpreted without knowledge of the results of the index test?	Did all participants receive a reference standard?
	Did the study avoid inappropriate exclusions?			Did all participants receive the same reference standard?
				Were all participants included in the analysis?
Risk of bias: High/low/unclear	Could the selection of participants have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct, or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Concerns regarding applicability: High/low/unclear	Are there concerns that the included participants do not match the review question?	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Are there concerns that the target condition as defined by the reference standard does not match the review question?	

Appendix 4. Anchoring statements for quality assessment of CSF tau and tau/ABeta ratio biomarkers diagnostic studies

Category	Review question	Inclusion criteria
Participants	Participants with mild cognitive impairment, no dementia	Participants fulfilling the criteria for the clinical diagnosis of MCI at baseline
Index test	CSF t-tau; CSF p-tau; CSF t-tau/ABeta ratio	CSF t-tau; CSF p-tau; CSF t-tau/ABeta ratio
Target condition	Alzheimer's disease dementia (conversion from MCI to Alzheimer's disease dementia) Any other forms of dementia (conversion from MCI to any other forms of dementia)	Alzheimer's disease dementia (conversion from MCI to Alzheimer's disease dementia) Any other forms of dementia (conversion from MCI to any other forms of dementia)
Reference standard	NINCDS-ADRDA; DSM; ICD; McKeith criteria; Lund criteria; NINDS-ARIEN criteria	NINCDS-ADRDA; DSM; ICD; McKeith criteria; Lund criteria; NINDS-ARIEN criteria
Outcome	N/A	Data to construct 2 X 2 table
Study design	N/A	Longitudinal cohort studies and nested case-control studies if they incorporate a delayed verification design (case-control nested in cohort studies)

Anchoring statements for quality assessment CSF tau and CSF tau/ABeta ratio

We provide some core anchoring statements for quality assessment of diagnostic test accuracy review of CSF tau and CSF tau/ABeta ratio biomarkers in dementia. These statements are designed for use with the QUADAS-2 tool and are based on the guidance for quality assessment of diagnostic test accuracy reviews of IQCODE in dementia (Quinn 2014).

During a two-day, multidisciplinary focus group and the piloting/validation of the guidance, it was clear that certain issues were key to assessing quality, while other issues were important to record but less important for assessing overall quality. To assist, we describe a 'weighting' system. Where an item is weighted 'high risk' then that section of the QUADAS-2 results table is likely to be scored as 'high risk of bias'. For example, in dementia diagnostic test accuracy studies, ensuring that clinicians performing dementia assessment are blinded to results of index test is fundamental. If this blinding was not present, then the item on the reference standard should be scored 'high risk of bias', regardless of the other contributory elements.

In assessing individual items, the score of 'unclear' should only be given if there is genuine uncertainty. In these situations, review authors will contact the relevant study teams for additional information.

Anchoring statements to assist with assessment for risk of bias

Patient selection

Was the sampling method appropriate?

Where sampling is used, the designs least likely to cause bias are consecutive sampling or random sampling. Sampling that is based on volunteers or selecting subjects from a clinic or research resource is prone to bias.

Weighting: High risk of bias ('no')

Was a case-control or similar design avoided?

Designs similar to case-control that may introduce bias are those designs in which the study team deliberately increase or decrease the proportion of subjects with the target condition, which may not be representative. For example, a population study may be enriched with extra dementia subjects from a secondary care setting, who are typically more diseased. Some case-control methods may already be excluded if they mix subjects from various settings.
Weighting: High risk of bias ('no')
Are exclusion criteria described and appropriate?
The study will be automatically graded as unclear if exclusions are not detailed (pending contact with study authors). Where exclusions are detailed, the study will be graded as 'low risk' if exclusions are felt to be appropriate by the review authors. Certain exclusions common to many studies of dementia are: medical instability; terminal disease; alcohol/substance misuse; concomitant psychiatric diagnosis; other neurodegenerative condition. Exclusions are not felt to be appropriate if 'difficult to diagnose' participants are excluded.
Post hoc and inappropriate exclusions will be labelled 'high risk' of bias.
Weighting: High risk ('no')

Index test

Was CSF tau and CSF tau/ABeta ratio biomarkers' assessment/interpretation performed without knowledge of clinical dementia diagnosis?
Terms such as 'blinded' or 'independently and without knowledge of' are sufficient and full details of the blinding procedure are not required. Interpretation of the results of the index test may be influenced by knowledge of the results of reference standard. If the index test is always interpreted prior to the reference standard, then the person interpreting the index test cannot be aware of the results of the reference standard and so this item could be rated as 'yes'.
For certain index tests the result is objective and knowledge of reference standard should not influence result, for example level of protein in cerebrospinal fluid, in this instance the quality assessment may be 'low risk' even if blinding was not achieved.
Weighting: High risk ('no')
Were CSF tau and CSF tau/ABeta ratio biomarkers' thresholds prespecified?
For scales and biomarkers there is often a reference point (in units or categories) above which subjects are classified as 'test positive'; this may be referred to as threshold; clinical cutoff or dichotomisation point. A study is classified 'high risk of bias' if the authors define the optimal cut-off post-hoc based on their own study data because selecting the threshold to maximise sensitivity and/specificity may lead to overoptimistic measures of test performance.
Certain papers may use an alternative methodology for analysis that does not use thresholds and these papers should be classified as not applicable.
Weighting: High risk ('no')

Reference standard

Is the assessment used for clinical diagnosis of dementia acceptable?
Commonly used international criteria to assist with clinical diagnosis of dementia include those detailed in DSM-IV and ICD-10. Criteria specific to dementia subtypes include but are not limited to NINCDS-ADRDA criteria for Alzheimer's dementia; McKeith criteria for Lewy Body dementia; Lund criteria for frontotemporal dementias; and the NINDS-AIREN criteria for vascular dementia. Where the criteria used for assessment is not familiar to the review authors or the Cochrane Dementia and Cognitive Improvement group ('unclear'), this item should be classified as 'high risk of bias'.
Weighting: High risk ('no')
Was clinical assessment for dementia performed without knowledge of the CSF tau and CSF tau/ABeta ratio biomarkers?
Terms such as 'blinded' or 'independently and without knowledge of' are sufficient and full details of the blinding procedure are not required. Interpretation of the results of the reference standard may be influenced by knowledge of the results of index test.
Weighting: High risk ('no')

Patient flow

Was there an appropriate interval between CSF tau and CSF tau/ABeta ratio biomarkers and clinical dementia assessment?
*As we test the accuracy of the CSF tau and CSF tau/ABeta ratio biomarkers for MCI conversion to dementia, there will always be a delay between the index test and the reference standard assessments. The time between reference standard and index test will influence the accuracy (*Geslani 2005*;* Okello 2009*;* Visser 2006*), and therefore we will note time as a separate variable (both within and between studies) and will test its influence on the diagnostic accuracy. We have set a minimum mean time to follow-up assessment of one year. If more than 16% of subjects of subjects have assessment for MCI conversion before nine months this item will score 'no'.*
Weighting: High risk ('no')

Did all subjects get the same assessment for dementia regardless of CSF tau and CSF tau/ABeta ratio biomarkers?
There may be scenarios where subjects who score 'test positive' on index test have a more detailed assessment. Where dementia assessment differs between subjects, this should be classified as 'high risk of bias'.
Weighting: High risk ('no')
Were all participants who received CSF tau and CSF tau/ABeta ratio biomarkers' assessment included in the final analysis?
If the number of participants enrolled differs from the number of participants included in the 2 X 2*table, then there is the potential for bias. If participants lost to dropout differ systematically from those who remain, then estimates of test performance may differ.*
If dropouts, these should be accounted for; a maximum proportion of dropouts to remain 'low risk of bias' has been specified as 20%.
Weighting: High risk ('no')
Were missing or uninterpretable CSF tau and CSF tau/ABeta ratio biomarkers results reported?
Where missing or uninterpretable results are reported, and if there is substantial attrition (we have set an arbitrary value of 50% missing data); this should be scored as 'no'. If those results are not reported, this should be scored as 'unclear' and authors will be contacted.
Weighting: High risk ('no' and 'unclear')

Anchoring statements to assist with assessment for applicability

Patient selection
Were included participants representative of the general population of interest?
The included participants should match the intended population as described in the review question. The review authors should consider population in terms of: symptoms; pretesting; potential disease prevalence; setting.
If there is a clear ground for suspecting an unrepresentative spectrum the item should be rated 'poor applicability'.
Index test
Were sufficient data on CSF tau and CSF tau/ABeta ratio biomarkers' application given for the test to be repeated in an independent study?
*Variation in technology, test execution, and test interpretation may affect estimate of accuracy. In addition, the background, and training/ expertise of the assessor should be reported and taken in consideration. If CSF tau and CSF tau/ABeta ratio biomarkers w*ere *not performed consistently, this item should be rated 'poor applicability'.*
Reference standard
Was clinical diagnosis of dementia made in a manner similar to current clinical practice?
For many reviews, inclusion criteria and assessment for risk of bias will already have assessed the dementia diagnosis. For certain reviews, an applicability statement relating to reference standard may not be applicable. There is the possibility that a form of dementia assessment, although valid, may diagnose a far larger proportion of subjects with disease than usual clinical practice. In this instance, the item should be rated 'poor applicability'.

CONTRIBUTIONS OF AUTHORS

CWR: overall responsibility of study design; overall responsibility of study selection and data extraction; advised about data and analyses; drafted Discussion section and finalised manuscript.

NS: designed and drafted protocol; study selection and data extraction; completed characteristics of included and excluded studies tables; entered data and data entry check; QUADAS-2 assessment; set up data and analysis tables; drafted analysis plan; completed SOF table and additional tables; updated Methods and drafted the Results and Discussion sections; finalised manuscript; managed the review process and produced progress reports, attended progress meetings and worked with all review authors to ensure that the review met publication deadlines.

EL: designed and drafted protocol; initial screening and study selection.

ANS: wrote the search strategy, searched and undertook initial screening of search results; study selection and data extraction; data entry check; QUADAS-2 assessment.

OU: performed statistical analyses and reviewed the draft manuscript.

SM: study selection; managed database; contributed to the Findings and Discussion sections

DECLARATIONS OF INTEREST

Craig Ritchie - None known

Nadja Smailagic - None known

Anna H Noel-Storr - None known

Obioha Ukoumunne - None known

Emma C Ladds - None known

Steven Martin - None known

SOURCES OF SUPPORT

Internal sources

- No sources of support supplied

External sources

- NIHR, UK.

This review was supported by the National Institute for Health Research (NIHR), via a Cochrane Programme Grant to the Cochrane Dementia and Cognitive Improvement group. The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the Systematic Reviews Programme, NIHR, National Health Service or the Department of Health

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

In our published protocol (Ritchie 2011), we stated that the minimum period of delay in the verification of the diagnosis (i.e. the time between the assessment at which a diagnosis of MCI is made and the assessment at which the diagnosis of dementia is made) was one year. In the preparation of the review, this criterion was not followed and we did not put limits on the length of duration of follow-up.

With respect to Investigation of heterogeneity, we planned to formally investigate the following, but these assessments of the sources of heterogeneity were not undertaken:

- Criteria used for definition of cognitive impairment
- Reference standards
- Participant sampling
- Index tests methodology used
- Duration of follow up and

We also planned to perform a sensitivity analysis for the individual quality items., but we were not able to do it, due to the small number of studies included.

INDEX TERMS

Medical Subject Headings (MeSH)

Alzheimer Disease [diagnosis]; Amyloid beta-Peptides [*cerebrospinal fluid]; Biomarkers [cerebrospinal fluid]; Cognition Disorders [diagnosis]; Sensitivity and Specificity; tau Proteins [*cerebrospinal fluid]

MeSH check words

Aged; Humans; Middle Aged